

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2004, 03:55:28 ; Search time 134.581 Seconds  
(without alignments)  
1613.049 Million cell updates/sec

Title: US-10-691-383-2

Perfect score: 3528

Sequence: 1 MLCHAADTRTGSMPDGTGL.....SIDGDMCSGLVYGVADCOA 676

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1342398 segs, 321133274 residues

Total number of hits satisfying chosen parameters: 1342398

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3528	100.0	676	9	US-09-840-762A-2
2	3528	100.0	676	16	US-10-691-383-2
3	130.5	3.7	3192	14	US-10-132-134-10
4	127.5	3.6	650	15	US-10-369-493-18669
5	123.5	3.5	3798	13	US-10-014-717-6
6	123	3.5	1939	12	US-10-152-886-43
7	122	3.5	22	9	US-09-840-762A-9
8	122	3.5	22	16	US-10-691-383-9
9	120.5	3.4	6310	12	US-10-282-122A-67793
10	117	3.3	888	14	US-10-156-761-8791
11	117	3.3	1387	16	US-10-437-963-167961
12	116	3.3	5245	14	US-10-329-079-11
13	114.5	3.2	912	9	US-09-746-491-49
14	114	3.2	2365	14	US-10-156-761-14890
15	113.5	3.2	912	9	US-09-905-983-2

16	113.5	3.2	3472	13	US-10-027-806-4	Sequence 4, Appli
17	113.5	3.2	3472	13	US-10-034-623-4	Sequence 4, Appli
18	113.5	3.2	3472	14	US-10-027-801-4	Sequence 4, Appli
19	113.5	3.2	3472	14	US-10-029-120-4	Sequence 4, Appli
20	113	3.2	4999	9	US-09-976-059-15	Sequence 15, Appl
21	112.5	3.2	730	16	US-10-437-963-122066	Sequence 122066,
22	111.5	3.2	1385	14	US-10-169-297-46	Sequence 46, Appl
23	111.5	3.2	1387	14	US-10-169-297-14	Sequence 14, Appl
24	111	3.1	391	16	US-10-437-963-132036	Sequence 132036,
25	110	3.1	21	16	US-10-691-383-11	Sequence 11, Appl
26	109.5	3.1	453	14	US-10-156-761-8288	Sequence 8288, Ap
27	109.5	3.1	1955	15	US-10-369-493-2646	Sequence 2646, Ap
28	109	3.1	611	15	US-10-369-493-12809	Sequence 12809, A
29	108.5	3.1	165	14	US-10-080-170-629	Sequence 629, App
30	108.5	3.1	165	16	US-10-080-170-629	Sequence 629, App
31	108.5	3.1	1439	12	US-10-282-122A-44567	Sequence 44567, A
32	108	3.1	19	9	US-09-840-762A-10	Sequence 10, Appl
33	108	3.1	19	16	US-10-691-383-10	Sequence 10, Appl
34	107.5	3.0	637	12	US-10-282-122A-69863	Sequence 69863, A
35	107.5	3.0	2394	14	US-10-329-079-47	Sequence 47, Appl
36	107	3.0	1273	12	US-10-363-616-325	Sequence 325, App
37	107	3.0	2273	12	US-10-282-122A-66115	Sequence 66115, A
38	107	3.0	2435	12	US-10-282-122A-47453	Sequence 47453, A
39	106.5	3.0	947	12	US-10-282-122A-63665	Sequence 63665, A
40	106	3.0	303	12	US-10-282-122A-69701	Sequence 69701, A
41	106	3.0	426	15	US-10-369-493-1118	Sequence 1118, Ap
42	106	3.0	1084	16	US-10-437-963-152384	Sequence 152384,
43	106	3.0	3705	12	US-10-282-122A-77944	Sequence 77944, A
44	105.5	3.0	451	14	US-10-156-761-10605	Sequence 10605, A
45	105.5	3.0	632	15	US-10-369-493-15431	Sequence 15431, A

ALIGNMENTS

RESULT 1

US-09-840-762A-2  
; Sequence 2, Application US/09840762A  
; Patent No. US20020035245A1  
; GENERAL INFORMATION:  
; APPLICANT: Vreeland, Valerie  
; APPLICANT: Ng, Kwan L.  
; APPLICANT: The Regents of the University of California  
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses  
; FILE REFERENCE: 023070-08710005  
; CURRENT APPLICATION NUMBER: US/09/840,762A  
; CURRENT FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: 09/151,189  
; PRIOR FILING DATE: 2001-04-23  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 676  
; TYPE: PRT  
; ORGANISM: Fucus distichus  
; US-09-840-762A-2

Query Match	100.0%;	Score 3528;	DB 9;	Length 676;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 676;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MLCHAADTRTGSMPDGTGLRLTSEORAKGWRQLEGEKSLGPHSPETPYIKYLEGSET 60		
Db	1	MLCHAADTRTGSMPDGTGLRLTSEORAKGWRQLEGEKSLGPHSPETPYIKYLEGSET 60		
Qy	61	WKVKUPTDGISASKILGKIMARVATATAVLAAPCLAFDEVTASGVPEEHKHTEG 120		
Db	61	WKVKUPTDGISASKILGKIMARVATATAVLAAPCLAFDEVTASGVPEEHKHTEG 120		
Qy	121	RHLQTCNSDDADPTAPNRDNNVAFASRRDARRRDGTGTCOITNGTDLATMPHKS 180		
Db	121	RHLQTCNSDDADPTAPNRDNNVAFASRRDARRRDGTGTCOITNGTDLATMPHKS 180		

181 LPHELGQVTTADDPAILEDCLNGDFSI C E D V P A G D P A G R L V N P T A A F A I D I S G P A S A T 240  
181 LPHELGQVTTADDPAILEDCLNGDFSI C E D V P A G D P A G R L V N P T A A F A I D I S G P A S A T 240  
241 T I P P V P T L S S P E L A A Q A E L Y M W A L A R D V P F M Q Y G T D E I T T T A A N L A G M G G P P N L D A V S 300  
241 T I P P V P T L S S P E L A A Q A E L Y M W A L A R D V P F M Q Y G T D E I T T T A A N L A G M G G P P N L D A V S 300  
301 I G S D G T V D P S Q L F R A T F V G V E T G P F V S Q L L V N S F T I D A I T V E P K Q T F A P D L N Y M W D F D 360  
301 I G S D G T V D P S Q L F R A T F V G V E T G P F V S Q L L V N S F T I D A I T V E P K Q T F A P D L N Y M W D F D 360  
361 E W L N I Q N G G P P A G P E E L D E L R F I R N A R D L A R V S F V D N I N T E A Y R G S L I L L E L G A F S R P G 420  
361 E W L N I Q N G G P P A G P E E L D E L R F I R N A R D L A R V S F V D N I N T E A Y R G S L I L L E L G A F S R P G 420  
421 I N G P F I D S D R Q A G F V N F G T S H Y F R L I G A A E L A Q R A S C Y Q K W Q V H R F A R P E A L G T L H N T I 480  
421 I N G P F I D S D R Q A G F V N F G T S H Y F R L I G A A E L A Q R A S C Y Q K W Q V H R F A R P E A L G T L H N T I 480  
481 A G D L D A D F D I S L L E N D E L L K R V A E I N A A Q N P N N E V T Y L L P Q A I Q V G S P T H P S Y P S G H A T Q 540  
481 A G D L D A D F D I S L L E N D E L L K R V A E I N A A Q N P N N E V T Y L L P Q A I Q V G S P T H P S Y P S G H A T Q 540  
541 N G A F A T V L K A L I G L D R G G E C F P N P V P S D D G L E L I N F E G A C L T Y E G I N K L A V N A F G R Q 600  
541 N G A F A T V L K A L I G L D R G G E C F P N P V P S D D G L E L I N F E G A C L T Y E G I N K L A V N A F G R Q 600  
601 M L G I H Y R F D G I Q G L L G E T T I V T R L H Q E L M T F A E A T F E F R L T G E V I K L F Q D G T F S I D G 660  
601 M L G I H Y R F D G I Q G L L G E T T I V T R L H Q E L M T F A E A T F E F R L T G E V I K L F Q D G T F S I D G 660  
661 D M C S G L V Y T G V A D C Q A 676  
661 D M C S G L V Y T G V A D C Q A 676

## RESULT 2

US-10-691-383-2  
; Sequence 2, Application US/10691383  
; Publication No. US20040110260A1  
; GENERAL INFORMATION:  
; APPLICANT: Vreeland, Valerie  
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases  
; TITLE OF INVENTION: and Their Uses  
; FILE REFERENCE: 023070-0871110US  
; CURRENT FILING DATE: 2003-10-21  
; PRIOR FILING DATE: 2000-06-19  
; PRIOR FILING DATE: 1998-09-10  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 676  
; TYPE: PRT  
; ORGANISM: Fucus distichus  
US-10-691-383-2

Query Match 100.0%; Score 3528; DB 16; Length 676;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 676; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLCHAADTRGSPMPDGVRLTLTSEQRAGWRRLQEGEKSGLGPHSPSETPIYKYLEGSET 60  
Db 1 MLCHAADTRGSPMPDGVRLTLTSEQRAGWRRLQEGEKSGLGPHSPSETPIYKYLEGSET 60  
Qy 61 WKVKLPDGTGISASKILGKIMARVRIATALAVLAAPCLAFDEVATAGVPEEHKHTGEG 120  
Db 61 WKVKLPDGTGISASKILGKIMARVRIATALAVLAAPCLAFDEVATAGVPEEHKHTGEG 120

121 RHLQTCNSDDALDPTAPNRDNVAFASRRDRAARRRDGTGTVCQITNGETDLATMEPHKS 180  
121 RHLQTCNSDDALDPTAPNRDNVAFASRRDRAARRRDGTGTVCQITNGETDLATMEPHKS 180  
181 LPHELGQVTTADDPAILEDCLNGDFSI C E D V P A G D P A G R L V N P T A A F A I D I S G P A S A T 240  
181 LPHELGQVTTADDPAILEDCLNGDFSI C E D V P A G D P A G R L V N P T A A F A I D I S G P A S A T 240  
241 T I P P V P T L S S P E L A A Q A E L Y M W A L A R D V P F M Q Y G T D E I T T T A A N L A G M G G P P N L D A V S 300  
241 T I P P V P T L S S P E L A A Q A E L Y M W A L A R D V P F M Q Y G T D E I T T T A A N L A G M G G P P N L D A V S 300  
301 I G S D G T V D P S Q L F R A T F V G V E T G P F V S Q L L V N S F T I D A I T V E P K Q T F A P D L N Y M W D F D 360  
301 I G S D G T V D P S Q L F R A T F V G V E T G P F V S Q L L V N S F T I D A I T V E P K Q T F A P D L N Y M W D F D 360  
361 E W L N I Q N G G P P A G P E E L D E L R F I R N A R D L A R V S F V D N I N T E A Y R G S L I L L E L G A F S R P G 420  
361 E W L N I Q N G G P P A G P E E L D E L R F I R N A R D L A R V S F V D N I N T E A Y R G S L I L L E L G A F S R P G 420  
421 I N G P F I D S D R Q A G F V N F G T S H Y F R L I G A A E L A Q R A S C Y Q K W Q V H R F A R P E A L G T L H N T I 480  
421 I N G P F I D S D R Q A G F V N F G T S H Y F R L I G A A E L A Q R A S C Y Q K W Q V H R F A R P E A L G T L H N T I 480  
481 A G D L D A D F D I S L L E N D E L L K R V A E I N A A Q N P N N E V T Y L L P Q A I Q V G S P T H P S Y P S G H A T Q 540  
481 A G D L D A D F D I S L L E N D E L L K R V A E I N A A Q N P N N E V T Y L L P Q A I Q V G S P T H P S Y P S G H A T Q 540  
541 N G A F A T V L K A L I G L D R G G E C F P N P V P S D D G L E L I N F E G A C L T Y E G I N K L A V N A F G R Q 600  
541 N G A F A T V L K A L I G L D R G G E C F P N P V P S D D G L E L I N F E G A C L T Y E G I N K L A V N A F G R Q 600  
601 M L G I H Y R F D G I Q G L L G E T T I V T R L H Q E L M T F A E A T F E F R L T G E V I K L F Q D G T F S I D G 660  
601 M L G I H Y R F D G I Q G L L G E T T I V T R L H Q E L M T F A E A T F E F R L T G E V I K L F Q D G T F S I D G 660  
661 D M C S G L V Y T G V A D C Q A 676  
661 D M C S G L V Y T G V A D C Q A 676

## RESULT 3

US-10-132-134-10  
; Sequence 10, Application US/10132134  
; Publication No. US20030171562A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Yang, Xianshu  
; APPLICANT: Staffa, Alfredo  
; APPLICANT: Zazopoulos, Emmanuel  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES  
; FILE REFERENCE: 3012-2US  
; CURRENT APPLICATION NUMBER: US/10/132,134  
; CURRENT FILING DATE: 2002-04-26  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 10  
; LENGTH: 3192  
; TYPE: PRT  
; ORGANISM: Streptomyces platensis subsp. roseaceus  
US-10-132-134-10

Query Match 3.7%; Score 130.5; DB 14; Length 3192;  
Best Local Similarity 19.0%; Pred. No. 0.053;  
Matches 150; Conservative 84; Mismatches 254; Indels 303; Gaps 38;

Qy 5 AADTRGSPMPDTRGV-----LRLTSEQRAGWRRLQEGEKSIGFPH--- 45  
Db 2228 AADVPSAPAEASMPOMSSAPATLDRDDFYAALGTAGLDARPYARSVEGVTELDARLL 2287  
Qy 46 -----PS--ETPIKYKLEGSETWKVKLPDGTGIS-----ASKILGKIMARVRI 86  
Db 2288 VRVAEPAMCQDPKHQVH-LPAWALVGL-TQGVQHWAGRAAIVRVVGVSGQEWETR- 2344

Qy 87 ATALAVVLAAPCLAFDEVTASGVFPEHKGHTG-----EGRHLOTCNTSD---DALDPTAPN 139  
Db 2345 ----AIVLAR-----TSDAVF-----HAAFLDEGRVLRGVEDAEFTAGDLEPALPG 2387  
Qy 140 RRDVAF-----SRRDARERDGTGTVCQITNG----- 169  
Db 2388 EAGRALVALPOASRPVLETPVGTGEMOQSEAVRPEAPSVTVAAVADGPAALVASLRETV 2447  
Qy 170 -----ETDLATMPHK-SLPDELGOVTADDFAILEDCLINGDFSICEDVPAGDPAG 219  
Db 2448 ADLLKFDLADLDLTHHAGVFESIALAKASELNGVLGTDLTPAVFFESD----- 2499  
Qy 220 RLNVNPTAAFAIDISGPAFSATT-----IPVPVPTLSSPELAQAELAELYMALARDVPFMQY 274  
Db 2500 --IRSLAEYLDRYGPPELSPTSADAPAPVAATRSP-----VPMAP 2540  
Qy 275 GTDBITTTAAANLAGMGFPNLDVSGDGTVDPPSOLFR-----ATFVG-VETGPF 326  
Db 2541 GPDD---DAVAIVGAAGRFPDAD-----DLDTFWQLRAGBDLLIADYPGDRFDGPGY 2589  
Qy 327 VSQLLVNSFTIDAITVPEKQETAPDLNVMVDPE-----MLNQNGG--P 370  
Db 2590 AEVVARADFPKAGRIE-GVDRFDADFFHLRLEAEMLDPQHLALAEVTVMALENGYP 2648  
Qy 371 PAGPEEL-----DEBLRFTIRNARDLA-----RVSFVDNIN----- 400  
Db 2649 ARUPENTGVYGVSGSDYHLLNAGVAPDGTATGNASHMLANRISVYLDVHGSPSPVD 2708  
Qy 401 -----TEAYR-----GSLILLELGAAPRPGINGPFISSDROAGFVNF 437  
Db 2709 TACSSSLVALHRAVEHTRISRCENMAIAGGVNLLSVDTFAATHMAG-MLSPDGRCKTFSA 2767  
Qy 438 GTSHYFRLIG-AAELAQASCYOK-----MOWHFRAR----- 468  
Db 2768 GADGYRSEGVAAVLLKPLAQOARDGDAIWGVVGRGSAENHGGRAGSTAPNGKAAALIQ 2827  
Qy 469 -----PEALGGT-LHNTTAGDLDADFDISLENDLKKRVAEINAAQNPNNEVTYLLP 520  
Db 2828 DAMRGIDPDSIGYVEAHGTGTG-LGDPVEVNALDSAYRALTAEE----- 2870  
Qy 521 QAIQVGSPTHSYPS-----GHATQNGAFATVVKALIGLDRGECFP-----NP 564  
Db 2871 -----GGPFAARPCALGSVKNITNGHAESAAGLAGVLKVLAM-RHRELPPALHCDRLNP 2924  
Qy 565 VFPSDDGLELI 575  
Db 2925 HLPLDGGFEV 2935

RESULT 4  
US-10-369-493-18669  
; Sequence 18669, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 18669  
; LENGTH: 650  
; TYPE: PRT  
; ORGANISM: Halobacterium sp. NRC-1  
US-10-369-493-18669

Query Match 3.6%; Score 127.5; DB 15; Length 650;  
Best Local Similarity 22.6%; Pred. No. 0.007;  
Matches 97; Conservative 48; Mismatches 176; Indels 109; Gaps 22;  
Qy 10 RGSFPMPTGVRLRLTSEORAKGW-----RRQLEGEKSLGFHP-SETPYIKYLEGSETWK 62  
Db 237 RFGPRDRGDAPAVTSDSRHLSFLPLAHVLERLSGH-----FLPLASGAQVCYAESPDTLR 292  
Qy 63 K--VKLPTDGIKASKILGKIMARVRIATALLAVVLAAPCLAFDEVTASGVFPEHKGHTG 119  
Db 293 EDFGLFEPTSATSUPRVEKLYDTIREQAA-----DSPVSEVPF-EWATRVGR 339  
Qy 120 GRHLQTCNTSDDALDPTAPNR-RDNVAFASRRDAARRERD-----GTGTVCQITNGETDLA 174  
Db 340 AAH-----TTDDPGIGLRAAHALADTLVFSDIRDALGGEIEFFISGGSL-----SPELC 389  
Qy 175 TMFH-KSLPHDELGOVTADDFAILEDCLINGDFSICEDVPAGDPAGRLVNPTAAFAIDIS 233  
Db 390 ALFHGMGLP-----ILRG-YGLTETSPV-----LAVNPYEDPIVGTI 425  
Qy 234 GRAFSATTIPPVPTLSSPE-----LAAQLAELYMALARDVPFMQYGTDEI 279  
Db 426 GPPVTDTELTVTETIASPEQRQCDGAGBELLARGPQVFGYW-----GLPDA 473  
Qy 280 TTTAAANLAGMGFPNLDVSGDGTV---DPFSQLFR-ATFVGVTGPFVSQLLVNSF 335  
Db 474 TDAAFVTREGKEWFRGTGDVVELRPDGVYRFLERAKQLTLSTGKNVAPGPIEDAFAPASPL 533  
Qy 336 TIDAITVPEKQETAPDLNVMVDPD-----EMLNQNGGPPAGPELDELRLRINARDLAR 392  
Db 534 VAQAMVVGDCQKFSAIL--VPNFDVAVSEWAAASQETALP-----DDRDAICRDERVRAR 585  
Qy 393 V-SFVDNINT 401  
Db 586 IQSAVDDVNT 595

RESULT 5  
US-10-014-717-6  
; Sequence 6, Application US/10014717  
; Publication No. US20020192778A1  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Rose  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/10/014,717  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US/09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 3798  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-10-014-717-6

Query Match 3.5%; Score 123.5; DB 13; Length 3798;  
Best Local Similarity 21.9%; Pred. No. 0.35; Mismatches 185; Indels 153; Gaps 22;  
Matches 109; Conservative 51; Mismatches 185; Indels 153; Gaps 22;  
Qy 158 DGTGTVCQITNGETDLATMPFKSLPHDELGOVTADDFAILEDCL----- 200  
Db 2760 DGEHFVSSQPLPEPDLAALV-----EEAGRVFADLPVLFWCKFAGERLADVLTKTLA 2813  
Qy 201 --IL--NGDFSICEDVPAGDP-----AGRLVNPTAAFAIDISGPAFSATT 241

Db 2814 LEILPFGSGFDMABRIYRDSPIARYSGNIVRGVVEAARVAVVPSGMPFISILEIGAGTGATT 2873  
Qy 242 IPPVPTLSSPELAQLAELYMALA-----RDVPFMOYGTDEITTTTAAANLAGMGGFN 295  
Db 2874 AAVLPVLLPORTEYHFTDVSPFLFARAEORFRDYFLKYGLDVOQSPAGQGYAHORFDV 2933  
Qy 296 LDAVSGSDGTVDPSQLFRATFVGTVGTFPVFSOLLVNSFTIDAITVEPKQETPAP-DLN 354  
Db 2934 IVAANV-IHATRD-----IRAT-----AKELL-----SLLAPGGLL 2963  
Qy 355 YMVD---FDEWLNTQNGPPAGPELDELRF---IRNARD---LARVSFVNI----- 399  
Db 2964 VLVEGTGHPWFWDITT---GLIEGWOKYEDDLRIDHPLLPARTCDVLRVGFADAVSLPGD 3022  
Qy 400 -NTEAYGSLILLELGAFAFRSGINGPFIDSDRQAGFVNFCTSHVFRLLIGAEAAQASCY 458  
Db 3023 GSPAGILGQHVLS-----RAGTIAGAACDSSGES-----ATESPAARAVR 3063  
Qy 459 QKWQ-----VHRFA-----RPEALGGTLHNTIAGDL-----DADFDI 490  
Db 3064 QEWADGSADVHRWALERMVYFHRPPGQVWVHGELRTGGAFKALAGDLLLPEDTGOVV 3123  
Qy 491 SLENDELLKRVAEINAAQNNEVTVLLPQAIQVGSPTHPSPVSGHAT-----QNG 542  
Db 3124 AEVQGLRLPQLEASAFAPRDPREEWLYALEWQRKDPPEAPAAAASSSSAGAWLVLMDOGG 3183  
Qy 543 AFATVLKALJLDRGGEC 560  
Db 3184 TGAALVSLLEG--RGEAC 3199

RESULT 6  
US-10-152-886-43  
; Sequence 43, Application US/10152886  
; Publication No. US20030064491A1  
; GENERAL INFORMATION:  
; APPLICANT: ECOPIA BIOSCIENCES INC.  
; APPLICANT: Farnet, Chris  
; APPLICANT: Staffs, Alfredo  
; APPLICANT: Zazopoulos, Emmanuel  
; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF ENEDIYNE RING  
; TITLE OF INVENTION: STRUCTURES  
; FILE REFERENCE: 3011-3US  
; CURRENT APPLICATION NUMBER: US/10/152,886  
; CURRENT FILING DATE: 2002-05-21  
; NUMBER OF SEQ ID NOS: 102  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 43  
; LENGTH: 1939  
; TYPE: PRT  
; ORGANISM: Amycolatopsis orientalis  
US-10-152-886-43

Query Match 3.5%; Score 123; DB 12; Length 1939;  
Best Local Similarity 22.6%; Pred. No. 0.13;  
Matches 146; Conservative 59; Mismatches 249; Indels 192; Gaps 32;  
Qy 4 HAADTRGSPMPDTGVRLITSEORAK---GWRQLEGEKSLGFHPSETPYIKVLESET 60  
Db 606 HADDIYRAANL-STGADQVATDVAQPRIVTGSGLAGLVKSLGIEATV--TGHSLGELT 662  
Qy 61 ---WKKVKLPTDGISASKILGKIMARV-----RIATALA----- 91  
Db 663 ALHWGGALTEREVVLKAKIRKQVWATASDGDGAWAATAATPSVAEGLAESEVVVIAGYNA 722  
Qy 92 ---VVLAAFLAFDEVTA-----SGVPEEHKHTGEGRLHQ--- 124  
Db 723 PEQTVLSPGAEDRVARARAEGVTAARINVSHAFSPAVVPAEAEMTGLAIDAIDPARL 782  
Qy 125 ----TCTNSDADLPTAPNR---RDNVAFASR-RDAARRRDGTGTCQITNGE----- 170  
Db 783 DRPVVSTVTGDLVHAAEDRLDRDQVVLVPRFREAAAKAERSDLVIEVGPRVLTGLL 842

Qy 171 -----TDLATMFKHSLPHDELQGVTTADDPFAI---LEDICIL-----NGDFSII-- 208  
Db 843 GTIAPGTPTVLSITDSTLTLPVLKVAAGAFAGQLESTLFDGRVVVRALPADGEFSFLA 902  
Qy 209 --CEDVP---AGDPAGRLVNPTAAFAIDISGPAPSAATTIPPVPTLSPELAQAELIYWM 263  
Db 903 SPCEAAPSIGAVLTRDRVAPBAEAAAGTASEGSGST---LDLLR--KLASERVELPLE 956  
Qy 264 ALARDV-PFMQYGTDEIT-----TTAAANLAGMGFPNLDVAVSIG-DGTVDPSOLF 314  
Db 957 AVTADTHPLDDLHLSITVQLVNDVTRALGRPALGMPNFATVCLGELAEIMIDELAOTA 1016  
Qy 315 R-----ATFVGVEGTGPFVSQLLVNSFTIDAITVEPKQETFAFDNLVWVDFEWLNTQNG 368  
Db 1017 KPADSQAQVAGV--GPW-----VRPFAVEVY-VAKP---SPDLATGISTAEWTAAPA 1065  
Qy 369 GPPAGPEELDEELFRIRNARDLARVFDVNDINTEAYRGSLILLELGAFAFRSGINGPFIDS 428  
Db 1066 GHP-----LAEPRLR-----AALATAGVGDG-----VLLCLNADSASGDVGLFLDA 1105  
Qy 429 DR-----QAGFVNEGTSHYFEL-----IGAAELAQASCYQKMQVHR 465  
Db 1106 GRAVLAAPNGTRFVVQVHGLGASGLAKTLRLDPSARTTIVDLADLGP----- 1153  
Qy 466 FARPEALGGTLHNTIAGDLDADDFDISLENDELLKRVAEINAAQNP 511  
Db 1154 -VDPEALDRAV-STVTEVAATTDSEVRVYDTAGVTRVTPKLAALTP 1197

RESULT 7  
US-09-840-762A-9  
; Sequence 9, Application US/09840762A  
; Patent No. US20020035245A1  
; GENERAL INFORMATION:  
; APPLICANT: Vreeland, Valerie  
; APPLICANT: Ng, Kwan L.  
; APPLICANT: The Regents of the University of California  
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses  
; FILE REFERENCE: 023070-087100US  
; CURRENT APPLICATION NUMBER: US/09/840,762A  
; CURRENT FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: 09/151,189  
; PRIOR FILING DATE: 2001-04-23  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: conserved  
; OTHER INFORMATION: vanadium-binding region 1, amino acids 452-473  
US-09-840-762A-9

Query Match 3.5%; Score 122; DB 9; Length 22;  
Best Local Similarity 100.0%; Pred. No. 7.9e-05;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 452 AQRASCYQKQVHRFARPEALG 473  
Db 1 AQRASCYQKQVHRFARPEALG 22

RESULT 8  
US-10-691-383-9  
; Sequence 9, Application US/10691383  
; Publication No. US20040110260A1  
; GENERAL INFORMATION:  
; APPLICANT: Vreeland, Valerie  
; APPLICANT: The Regents of the University of California  
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases  
; TITLE OF INVENTION: and Their Uses  
; FILE REFERENCE: 023070-087110US



3  
CURRENT APPLICATION NUMBER: US/10/691,383  
CURRENT FILING DATE: 2003-10-21  
PRIOR APPLICATION NUMBER: US/09/596,794  
PRIOR FILING DATE: 2000-06-19  
PRIOR APPLICATION NUMBER: US 09/151,189  
PRIOR FILING DATE: 1998-09-10  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 9  
LENGTH: 22  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: conserved Fucus  
OTHER INFORMATION: vanadium-binding region 1, amino acids 452-473,  
OTHER INFORMATION: 1st conserved motif  
US-10-691-383-9  
Query Match 3.5%; Score 122; DB 16; Length 22;  
Best Local Similarity 100.0%; Pred. No. 7.9e-05;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 452 AQRASYQKQVHRFARPEALG 473  
Db 1 AQRASYQKQVHRFARPEALG 22  
RESULT 9  
US-10-282-122A-67793  
Sequence 67793, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haseibeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA 034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 67793  
LENGTH: 6310

TYPE: PRT  
ORGANISM: Pseudomonas putida  
US-10-282-122A-67793  
Query Match 3.4%; Score 120.5; DB 12; Length 6310;  
Best Local Similarity 20.2%; Pred. No. 1.6;  
Matches 156; Conservative 271; Indels 279; Gaps 36;  
Qy 32 WRRQLEGEKSLGFHPSTPYIKYLEGSETWKKVLPDGDIGISASKILKIMARVRIATALA 91  
Db 86 WHAQYDAAAFGTFTFEVASLDELVAGIGTAGSAMPTWAIAGLSLLG-----VGGAAA 138  
Qy 92 VW-----LAAECLAFDEVTAGVPEEHKHTGEGHLOTCINSDA----- 132  
Db 139 AADNGGGSSGGSDPDTSAPATPIDLL-----VSPDGLRLTGRG-EAGTTVNIRDAAGNLI 193  
Qy 133 -----LQPTAPNRDRNVAFASSRRDARRERDGTGVCQITN 168  
Db 194 GSGTVGADGNFNTLNAPOINSENLOVTLTDAAGNV-----APGAVTAPDATAPEL----- 244  
Qy 169 GETDLA-----TMFHKSLPHDELQVTDADFAILEDCI--LNGDFSCIEDVDPAGDPAGR 220  
Db 245 APTDLAINEQGNLTGAEFGSTV-SVRGAGVLLGTAVAGADGQFSITLQPPQSD--GQ 301  
Qy 221 LVNPTAAFAIDISGPAPSATITIPPV--PTLSSPELAAQLAEIYWMALARVDFPMQYGTDE 278  
Db 302 ALBISAADAAGNISPVASITA-PDVNDPDTTAEQPTDL-----ALADGVFTTGRGEPG 354  
Qy 279 ITT-----TAAANLAGMGFPNLDVAVSGSDGTVPFSSOLFRATFVGTVETGPFVSOLLVNSF 335  
Db 355 ATVOVRDAAGNLITG-----VVGADG-----LFSLTLSPACT-----NGE 390  
Qy 336 TIDAITVEPKQETFAPDLNMYWDFD-----EWLNTQNGQPPAGPPELD 378  
Db 391 ALDVLVDAAGNSAP-----LQFDAPDITPPDAVTNITVGADGLALSGRGEPAATVB-- 443  
Qy 379 EELRFIRNARDLARVSVDNINTEAVRGSLLLELGAFSRPGINGPFIDSDRQAGFVNFG 438  
Db 444 -----VRDANGTVIGTVGVGAN-----GTFLIDLAPAAQGEQLSLVQTD-PSGNASVA 491  
Qy 439 TSHYFRLLIGRAE-----LAQRASCYQKQVHRFARPEALGGTLHTIAGDLD 485  
Db 492 TEYDVPLTTAPDSPSNLAIDADGTTLTGTAPAGSRVEVH-----DANGTLIGSAIANAD 545  
Qy 486 ADPDISI-----LENDELLKRYA-----EINAA 508  
Db 546 GSFSEIENPAQANGELLDVVAIDDDGGVSSLPQAQITAPDITAPAPTETELAVSADGSVITGR 605  
Qy 509 QNPNEVTVLLPQAIQVGS-----PT-----HPSYPSGHATQNGAFATVKKALIGLDRG 557  
Db 606 AEPGSTVRIVAADGTETLGTAVVGFTGVFSLNLPNPQVDGEVLQ--ATAT-----DAA 655  
Qy 558 GECFPNPVF--PSDDGLE-----LINFGACLTYEIGEINKLAVNVAFGQMLGIH 605  
Db 656 GNTSPSSAVTAPDIDGVDITPPAAPTDLVLIGAGSQLSGRGE----- 697  
Qy 606 YRFDGIQGLLGETITVTTLHQELMTFAEATFEFLFTGCEVIKLFODGTFSI 658  
Db 698 -----AGSTVQVRDAAGNI-----LATGTVA--ADGFTV 725  
RESULT 10  
US-10-156-761-8791  
Sequence 8791, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 8791  
; LENGTH: 888  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-8791

Query Match 3.3%; Score 117; DB 14; Length 888;  
Best Local Similarity 21.7%; Pred. No. 0.13;  
Matches 151; Conservative 85; Mismatches 255; Indels 204; Gaps 36;  
6 ADTRGSPMDTGVLRLTSEQ-----RAKGRERQLEGKS-----LGFHPSETP 50  
28 ADRASSVARPAQDVLSLQTERRLTAVQASRTKTARTELDGARDETDAVAAPRSSS- 86  
51 YIKYLEGSETWKVKVL---PTDGISASK--ILGKIMA-----RVRIATALAVVLAAP 97  
87 --SALDTSLSLQRRRLFDALDLSGRREALDGRTLSTGDTFELYTDTVSRGIGVLTAA- 143  
98 CLAFDEVTSAGVPPPEEHKHTGEBR-----HLQTCNSDDALDPTA--PNRR--- 141  
144 -----VHSDGRLARGGNATVSLAHITEMLSREDALISGALPSSRMATA 186  
142 -DNVAPASRRDAARRRDG-----TGTCQITNG-----ETDLATMPHKS 181  
187 TTRAQFOYLAIQREFRAGLDIHDLPAGAAATYQLTDSAKWTTLGTVKESVTTGRGTG 246  
182 PHDELQVTDADDFAILEDCLNGDFSCEDVP-----AGDPA-----GRLVNPTAAFAI 230  
247 P---GQ--ASSMPTATEPVVGGIQLGADSADGLSLDAADHADDLLGLMLG--TAATLA 299  
231 DISGPASFA-----TTTPPVPTLSS-----PELAQALAEIYMWALARDVPFMOYG 275  
300 ALAGGAVIALRARRSTLGRVSELQAHVEQLSGSWLPOLLARIQNGERVEPATLAPHGEQA 359  
276 TDEITTTAAANLAGMGFPNLDV--SIGSDGTVDPSQLFRATFV-----GVET 323  
360 TDELERIAAA-IDQLGRVAGDTAVRQSLGREGTEKVPQALIRRTQIILHRLISLLDLE 418  
324 GPFVSQLLVNSFTIDAITVEPKQFTFAPDLNMYVDFDEWLNIONGSGPPAGPEELBELRF 383  
419 KHEDSLLKDKIFKVDHLATVRRA-----ENLVLSGSP----- 454  
384 IRNARDLARVSFVNINTEAYRGLLLELGAERSRPGINGPFFIDSDRQAGFVNGTSHYF 443  
455 --SRRLTAPVSI-----TDVNRGA--VAETEQTTRVKVN--LPADRRLALAGRAVADVT 503  
444 RLIGAELAQRASYCQ--KQVHRFARPEALGGTLH-----NTIAGDLAD-- 487  
504 HLL--AELIENGTSFSPDQTFVSATKVAGLAHVHVEDHGLMPQDLRDLHANDLLAHP 561  
488 -FDISLLENDE-----LLKRVAEINAAQNPNNEVTV-----LLPOAI--QVGSPTHP 532  
562 RLDWTALGEDPRLGHFVVVARLAERHKIKVELRESVYGGTLVILVLLPAALLEEVASPVLDQ 621  
533 YPSGHATQNGAFAT-VLKALIGLDR-----GGECPF 562  
622 LKSAVAANRAVAEASRAVAGTDDFAVVGAEQMP 656

## RESULT 11

US-10-437-963-167961  
; Sequence 167961, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 167961  
; LENGTH: 1387  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_66522C.1.pap  
US-10-437-963-167961

Query Match 3.3%; Score 117; DB 16; Length 1387;  
Best Local Similarity 18.8%; Pred. No. 0.28;  
Matches 153; Conservative 89; Mismatches 273; Indels 298; Gaps 38;

26 EORAKWRQLEGKSLGPHSPSETPVIKYL--EGSETWKVKVLPDTGISAASKILGKI--- 80  
286 QSLEYLRKQAEYDVRG-----PVIDIVAHGDGVW-RVAVDTQGLEGNKCGKLADF 338  
81 --NARVRI-----ATALAVVLA--PCLAF 101  
339 VPLTNYRLERKFGIFSKLDACSFPVANIYDDGNLVSIVTDCSPHATHVAGIAAFAHP--- 394  
102 DEVTASGVPEEHKHTGEBRHLQTCNSDDALDPTAPNRDNVAFASRRDAARRDGTG 161  
395 DEPLLNGVAP-----GAQLISCKIGDTRLSGM-----ETGTG 426  
162 TV-----CQITN--GETDLATMPHKSIPHDELQVTDADFAI--LEDCLINGDF 206  
427 LVRALIAAVEHKCDLINMSYGEPTLLPDYGRFI---DLASEVVDKRIIFISAGNNGPA 483  
207 SICEDVPAGDPA-----GRLVNPTAA-----FAIDIS 233  
484 LNTVGAPGGTSSSIIGVAVVSPAMAAGAHCVVQAPAEQMEYTWSSRGPTADGDLGVSIS 543  
234 GPAFSATTTPPVPT-----LSPELAQALAEIYMWALARDVPFMOYGTDEIT 280  
544 AP---GGAVAPVPTWTLSRMLMNGTSMSSPSACGGVALLVSAKAEGLPSPTYVRKAI 600  
281 TTAANLAGM-----GGFPNLD-----AVSIGSDGTVDPSQLFR 315  
601 ENTAASISDVPEEKLTGHLQVDRAFEYAAQAKELPLVSYRISINQVG--KPTSKLRG 658  
316 ATVGVGVTGPFVSQLVNSFTIDAITVEPKQFTFAPDLNMYVDFDEWLNIONGSGPPAGPE 375  
659 IYLRGNTCQTSSEWTVQ-----LDPKFHEDASNMEOQVPEECQLHS--TDSSVI 708  
376 ELDEELRFRNARDLARVSFVDNINT-----EAY-----RGLLLELGAERSRPGIN 422  
709 KIPEYIMVTNNGRTFNIVNPNVNISSGLHYEYVYGDCKAPWRGPFPRVPTIKPIALS 768  
423 G--PFIDSDRQAGFVNGTSHYERL-----IGAA--ELAQRASYCQ----- 459  
769 GEPPALTLSN-----LSFKSGHIERFINVPDIGHASWVEVNTWRTSAFDTPRRFLDTVOICP 824  
460 -----KWQ-----VHRFARPEALGGTLHNTIAGDLDA-----DFDI----- 490  
825 LKRPKWEAVVTFSSPSLKNFSPVGGTLLELSIAQFWSGSIASHEPTCVDFEIVPHGI 884  
491 -----SLENDLKRVAEINAAQNPNNEVTVLLPQAIQVGSPT 529  
885 SVDQKIIGLDGSEAPVRVVARSLASERLVP-VATLNKVKTPYRPVESNLCSL-----PPS 939

Qy 530 HPSVPSGHATONGAPATVVKALIGLDRGCEFP-----NPVFPSSDDGLE-----LINPEG 579  
Db 940 RDRLPSPG---KQIIALTLYFKLEDAEIKRPVPLNNRIY--DNKFSEQYVIRSDSN 993  
Qy 580 ACLTYEGEINKLAVNVAFGROMLGIHYRFDGQI 612  
Db 994 KCVYSGDVVPYNYKLSKGEYTLQLYIRHDNVQ 1026

## RESULT 12

US-10-329-079-11  
; Sequence 11, Application US/10329079  
; Publication No. US20030198981A1  
; GENERAL INFORMATION:  
; APPLICANT: FARNET, Chris  
; APPLICANT: ZAZOPOULOS, Emmanuel  
; APPLICANT: STAFFA, Alfredo  
; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF LIPOPEPTIDES  
; FILE REFERENCE: 3002-11US  
; CURRENT APPLICATION NUMBER: US/10/329,079  
; CURRENT FILING DATE: 2002-12-24  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 11  
; LENGTH: 5245  
; TYPE: PRT  
; ORGANISM: Streptomyces fradiae  
US-10-329-079-11

Query Match 3.3%; Score 116; DB 14; Length 5245;  
Best Local Similarity 22.5%; Pred. No. 3.4;  
Matches 159; Conservative 66; Mismatches 233; Indels 248; Gaps 41;

Qy 13 PMPTGVLR-LTSEORAKGWRROLEGEKSLGFHPSET---PYKYLEGSETKVKVLP 68  
Db 3698 PRPE-----RLPLSHAQRRLMFVROLEG-----PSATYVNPWALRLTG-----PL 3737  
Qy 69 DGISASKILGIMARVRIATALAVVLAAPCLAFDEVASGVFPBEKHKTGEGHLOCTN 128  
Db 3738 DVAALRAALGDVVAR---HESLRIVFA-----EDEH-----GPH-QVVL 3773  
Qy 129 SDDALDPTAPNRRNVAF-----ASRRDAARRRGTVGVCOITNGETDLATMFH 178  
Db 3774 ADGPAPLSGVPRTDEDALPRLREADHAFRLDAEPPLR--AHLFATAPEDHTELLVMHH 3831  
Qy 179 KSLPHDELGO-----VTADDFAILE-----DCIL 202  
Db 3832 --IATDWSORPLTADLAAYAAHAGRVFTLPPLPVAYADYALWQOARLGDREKDSAL 3889  
Qy 203 NGDPSICEDVPAGDPAGRLVNPTAAFAIDISGPAFSATTIPPVPTLSSPELAAQLAEIYW 262  
Db 3890 SAQLAYWRDALAGSP-----EELALPADRPAPVPSHRGDSVPLTVPELHGEVVE--- 3940  
Qy 263 MALARD---VPFMQY-----GTDEITTTAAAN-----LAGMGGFNLDVSI 301  
Db 3941 --LAREFRATPFMVVHAALAAALLTRLGAGTDVPIGSPVAGRVDDALEDLVGF-FVNTLVL 3997  
Qy 302 GSDGTVP-PSOLF---RATFVGU---ETGPF-----VSQLLVNSFT- 336  
Db 3998 RTDTSGDPTFGELLERVRATDLGAYAHQDLUFERLVELRDPERSLARHPLFQVSLNYDTA 4057  
Qy 337 -----IDAITVEPK-----QETFPAD-----LNTMVD-FD 360  
Db 4058 ETARDAAPDELGLTVSGRPLGVTTSKFDLTFTALTETRAHDCGPGARGALEYSTDLFD 4117  
Qy 361 EWLNIQGGPPAGPEELDEELRFINARDLARVFDVNDINTEAYRGSLILLELGAFRPG 420  
Db 4118 R-----GTAERLAERPARVLQAAVAPGRLQIDV-LLPGRALLE-GEWSRPE 4165  
Qy 421 INGPFDSDRQAGVNF-----GTSHY-----FRLIGRAELAQASCYQKQVHRFAR 468  
Db 4166 -PGVAPTD-DARPPDLFEAQARTPHAPAVRDGDRELSVAELMDRANRLARFLAARGAG 4223

Qy 469 PEALGGTLHNTIAGLDLADDFDISLENDELLKRVAEINA-----AQNPNNEVTYLL 519  
Db 4224 PE-----DTVA-----VLLPRGPELITALTAVQKAGAAVVPMDAELPAERIAHML 4268  
Qy 520 ----PQAIQVSGSPHPSGSHATONGAPATVVK-ALIGLDRGCEC 560  
Db 4269 ENARFVLVLAHTATQDALPEGAGPVVRLDAPATEAALAGLD-GGDC 4313

## RESULT 13

US-09-746-491-49  
; Sequence 49, Application US/09746491  
; Patent No. US20020137202A1  
; GENERAL INFORMATION:  
; APPLICANT: Burgess, Catherine E.  
; TITLE OF INVENTION: No. US20020137202A1e1 Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 15966-621  
; CURRENT APPLICATION NUMBER: US/09/746,491  
; CURRENT FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: USSN 60/171,329  
; PRIOR FILING DATE: 1999-12-21  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 49  
; LENGTH: 912  
; TYPE: PRT  
; ORGANISM: Gallus gallus  
US-09-746-491-49

Query Match 3.2%; Score 114.5; DB 9; Length 912;  
Best Local Similarity 19.8%; Pred. No. 0.24;  
Matches 131; Conservative 82; Mismatches 232; Indels 215; Gaps 33;

Qy 107 SGVFPBEKHKTGGRHLQCTNSDDALDPTAPN-----RR 141  
Db 278 NGTVPEGSK---PQYVYMTVT-AIDADDPAQNAQMLRYILSQAPSSPSPMFTINNETG 333  
Qy 142 DNVAFASSRRDAAR-----RERDGTGT-----VCQIT-----NGETDLATMFH 178  
Db 334 DIITVAAGLDREKQVQVTLIIQATDMEGNPTYGLSNATATAVITVDNDNPPEETAMTFY 393  
Qy 179 KSLPHDELGOVTDADDFAILEDICILNGDFSCIEDVPAGDPAGR---LVNP-----TAAP 228  
Db 394 GEVPENRVDVIVANLTVTDKDPHTPAWARYQMTGGDTGQFTILTDPNSNDGLVTVVK 453  
Qy 229 AID-----ISGPAFSATTIP-----PVPTLSSPEL 253  
Db 454 PIDFETNRMFVLTVAAENQVPLAKGIQHPPOSTATVTSITVIDNESPYFVNPVKLVROEE 513  
Qy 254 AAQLAEIYMWALARDVP-FMQYGTDEITTTAAANLAGMGGFNLDVAVS--IGSDGTVDPPF 310  
Db 514 GLLAGSLMTTFTARDPPRYMQQ-----TSLRYSKLSDPANWLKIDPVNGQITTTAVLDRE 568  
Qy 311 S-----QLFRATFVGVEGT-PFVS-----QLLVNSFTIDAIVPEPKO-----ETAPDLNY 355  
Db 569 SIYVQNNMYNATFLASONGIPPMNSGTGLQIYLLIDINDNAPQVNPKEATTCTETLPNAIN 628  
Qy 356 MVDFDEWLNTQNGSP-----PAGEELDEELRFINARDLARVS----- 394  
Db 629 ITAVDPDID-PNAGPPAFAPFELPSPPSIKRNVITVRISGDHAQSLRIRFLEAGYIDVPIV 687  
Qy 395 FVDNINTEAYRGSLILLELGAFRSGPGPFIDSDRQAGFVNFQGTSHYFR-LIGAAELAQ 453  
Db 688 ITUSGNPHASSTSVLKVKV---CCQDINGDCTDVRIVG-AGLGTGAIILLCIILLI 743  
Qy 454 RASCYQKQVHRFARPEAL-----CGTLHNTIAGLDLADDFDISLL 493  
Db 744 LVLMEFVMMKRRDKERQAKQLLIDPEDDVRDNLKYEDEGG-----GEEQDYDLSQL 796  
Qy 494 ENDELLKRVAEINAQAQNPNNEVTYLLPQATQ-VG-----SPTH-----SGHATQ 540  
Db 797 QPDTVE-----PDAIKPVGIRLDERPIHAEPQYFVRSAAHPGPD 837

```
QY 541 NGAFATVKGALIGDRGCEFPNVPFSDGDLINFEAGACLTVEGINKLVNVAFRQ 600
Db 838 IGDF--INEGLKAADN-----DPTAPPYDLSLLVFDYEGSGST-AGSLSSLNSSSGGEQ 888

RESULT 14
US-10-156-761-14890
; Sequence 14890, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14890
; LENGTH: 2365
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14890

Query Match 3.2%; Score 114; DB 14; Length 2365;
Best Local Similarity 22.6%; Pred. No. 1.4;
Matches 142; Conservative 56; Mismatches 233; Indels 198; Gaps 31;

QY 57 GSETWKKVLPDGTGDISAKILGKIMARVRIATALLVLAAPCLAFDEVTASGVFPPEBKH 116
Db 1162 GARTW-----IALVAESAELALRLRR-AVAGEHAPGEGFTADDTE 1202

QY 117 TGE-----GRHLOTCTNSDDALDPTAPNRDRNVAFASRR 150
Db 1203 TGEVALLPFGGSGRPGMFAELFVAFPELQRLRLDETTARVLPF--PTAFDE--ASRK 1257

QY 151 DAARRRDGTGTCQITNGETDLATMFHKSILPHDELQVTTADDFAILEDCLINGDFICE 210
Db 1258 E--QOERITDTTVAQPALGLTGLAAF--QLLTRAGIRPANAAGHSYGELAALAAAGALTP 1313

QY 211 DVPAGDPAGRLVNPTAAFAIDISGPAFSATTIPVPPTLSSPELAALAEIYWMALARDVP 270
Db 1314 DALVRTSRGR-----AAVLGAAGDGPDTM--AAVAGEPEVAALAAA----- 1356

QY 271 FMOYGTDEITTTAAANLAGMGFFNLDAVSGSDGTVDPSQLFRATFVGVTGP-----F 326
Db 1357 -----GLDGSVVTAEN-----SPRQTVISGPTDDVLTAVERL-RAQGLGAKRIPVACAF 1405

QY 327 VSQLLVNSFTIDAITVEPKQTPAPDLNYM-----VDFEWMINTONGGPPAGPELDELR 382
Db 1406 HSPL-----VAAAGEAFAKVLADVPVGATDFPVMNSRNTAARYPESPEAIRAELA 1454

QY 383 FIRNARDLARVSFVDNINTEAYRGLSILLEGAFS-----RPGINGPFIDSRR 430
Db 1455 -----AQIGAPVRFPADQIEMYEAGARFVEAGPGSVLTRLVGTVGLDRPH-RTVALEKGR 1509

QY 431 QAGVNF-----GTSHYFRLIGAAELAQ-----RASCYQKQV----- 463
Db 1510 HTGLVGLAALAQALAVAGVDVRTSWLFGQRGAVDAARAPRVRAA-----VTVDGHLIRTA 1565

QY 464 -----HRFAR-PEALGGTGLNTTAGDLADDFDISLLENDELLKRVAEINAAQN--- 510
Db 1566 DGAIPAAALHPAERVPEAL--VTHSTSGAPVAATGSEALI--SEFLRTSREMIAAQRDVM 1621

QY 511 -----PNNEVTYLLPOAIOVGSPHTSPHSYPSGHAT-----QNGAF 544
Db 1622 LGYLGADEVPRPPAPAAAPATYADAPVTY-ADAPVTVTAP--PAAPAAVAAPAGGSV 1678

QY 545 ATVLKALIGLDRGCEFPNVPFSDGDL 573
Db 1679 LSVLDVIG-ERTG--YPVDMIEPDLDE 1704

RESULT 15
US-09-905-983-2
; Sequence 2, Application US/09905983
; Patent No. US20020045591A1
; GENERAL INFORMATION:
; APPLICANT: Geiger, Benjamin
; APPLICANT: Ben-Ze'ev, Avri
; APPLICANT: Sadoc, Einat
; TITLE OF INVENTION: METHODS AND THERAPEUTIC COMPOSITIONS FOR TREATING CANCER
; FILE REFERENCE: 01/22326
; CURRENT APPLICATION NUMBER: US/09/905,983
; CURRENT FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 912
; TYPE: PRT
; ORGANISM: Gallus gallus
US-09-905-983-2

Query Match 3.2%; Score 113.5; DB 9; Length 912;
Best Local Similarity 21.2%; Pred. No. 0.31;
Matches 113; Conservative 65; Mismatches 195; Indels 160; Gaps 28;

QY 175 TMFHKSILPH-----DELGOVTADDFAILEDCLINGDFESICEDVPAGPAGRLVNP----- 224
Db 409 TVTDXKQHPHTPAWNAHYQMTGGD-----PTGQFTILTDPNSDGLVTVVPIPETN 460

QY 225 -----TAA-----FAIDISGPAFSATTIP-----PVPTLSSPELAALAEI 260
Db 461 RMFVLTVAAENQVPLAKGHOHPQSTATVSIITVIDVNESYFVFNPKLVQEGSLLAGSM 520

QY 261 YWMALARDVP-FMOYGTDEITTTAAANLAGMGFFNLDAVS--IGSDGTVDPS-----Q 312
Db 521 LTTFTARDPRYMQQ-----TSLRYSKLSDPANWLKIDPVNGQITTTAVLDRESIYVQNN 575

QY 313 LFRATFVGVTG-PFVS-----QLLVNSFTIDAITVEPKQ-----ETFAPDLNVMVDFDEW 362
Db 576 MYNATFLASDNGIPPMPSGTGTQIYLLDINDNAPQVNPKEATTCTETLQPNAINITAVDPD 635

QY 363 LNIQNGGP-----PAGPELDEBLRFRNARDLARVS-----FVDNINT 401
Db 636 ID-PNAGPPAFAPELPDSPPSIKRWNTIVRISGDHAQLSLRIRFLEAGIYDVPVITDSGNP 694

QY 402 EAYRGLSILLELGAFAFRGNGINGPFIDSDROAGFVNFGTSHYFR-LIGAABELAQRASYQK 460
Db 695 HASSTSVLKVKV---CQCINGDCTDVRIVG-AGLGTGAIILLCIIILLILVLMFVV 750

QY 461 WQVHRFARPEAL-----CGTLHNTIAGDLADDFDISLLENDELLK 500
Db 751 WMKRRDKERQAKQLLIDPEDDVVRDNLKYEDEGG-----GBEDQDYDLSUQQQPTVE 803

QY 501 RVAEINAAQNPNNEVTYLLPOAIQ-VG-----SPTH--PSYP-----SGHATONGAFATV 547
Db 804 -----PDAIKPVGIRLDERPIHAEPQYVRSAAHPGIDGDF--I 842

QY 548 LKALIGLDRGCEFPNVPFSDGDLINFEAGACLTVEGINKLVNVAFRQ 600
Db 843 NEGLKAADN-----DPTAPPYDLSLLVFDYEGSGST-AGSLSSLNSSSGGEQ 888

Search completed: September 19, 2004, 04:03:35
Job time : 137.581 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2004, 03:52:28 ; Search time 29.3913 Seconds  
(without alignments)  
1187.397 Million cell updates/sec

Title: US-10-691-383-2  
Perfect score: 3528  
Sequence: 1 MLCXAADTTTGGSPMDTGV.....SIDGDMCGLVYTGADCOA 676

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PTUS COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3528	100.0	676	3	US-09-151-189-2
2	3528	100.0	676	4	US-09-596-794-2
3	129.5	3.7	3892	4	US-09-328-352-5503
4	123.5	3.5	2890	3	US-09-413-814-67
5	123.5	3.5	3798	3	US-09-335-409-6
6	123.5	3.5	3798	4	US-09-568-102-6
7	123.5	3.5	3798	4	US-09-567-969-6
8	123.5	3.5	3798	4	US-09-568-480-6
9	123.5	3.5	3798	4	US-09-568-486-6
10	123.5	3.5	3798	4	US-09-568-472-6
11	123.5	3.5	3798	4	US-09-567-899-6
12	122	3.5	22	3	US-09-151-189-9
13	122	3.5	22	4	US-09-596-794-9
14	117	3.3	1043	4	US-09-552-991A-32210
15	114	3.2	913	2	US-08-474-067-6
16	114	3.2	913	2	US-08-474-068A-6
17	114	3.2	913	2	US-08-472-481-5
18	113.5	3.2	3472	4	US-09-408-020-4
19	110	3.1	21	4	US-09-596-794-11
20	108.5	3.1	2315	4	US-09-543-681A-5434
21	108	3.1	19	3	US-09-151-189-10
22	108	3.1	19	3	US-09-596-794-10
23	102	2.9	440	4	US-09-684-855-101
24	102	2.9	440	4	US-09-684-855-124
25	102	2.9	440	4	US-09-684-855-146
26	102	2.9	466	3	US-08-868-435-2
27	102	2.9	466	3	US-08-868-435-12

28	102	2.9	466	4	US-08-744-231-2	Sequence 2, Appli
29	102	2.9	466	4	US-08-744-231-12	Sequence 12, Appli
30	102	2.9	466	4	US-09-636-499-7	Sequence 7, Appli
31	102	2.9	466	4	US-09-273-871A-12	Sequence 12, Appli
32	102	2.9	683	4	US-09-252-991A-33038	Sequence 33038, A
33	102	2.9	737	3	US-08-989-385-3	Sequence 3, Appli
34	102	2.9	737	4	US-09-593-826-3	Sequence 3, Appli
35	101.5	2.9	600	2	US-08-679-405-2	Sequence 2, Appli
36	101.5	2.9	600	3	US-08-842-799-2	Sequence 2, Appli
37	101.5	2.9	600	3	US-09-271-778-2	Sequence 2, Appli
38	101.5	2.9	600	4	US-09-788-871-2	Sequence 2, Appli
39	101.5	2.9	600	5	PCT-US96-11458-2	Sequence 2, Appli
40	101.5	2.9	654	4	US-09-543-681A-5710	Sequence 5710, Ap
41	100.5	2.8	564	4	US-09-252-991A-27329	Sequence 27329, A
42	100.5	2.8	1042	3	US-08-928-361B-11	Sequence 11, Appl
43	100.5	2.8	1042	4	US-09-588-995A-11	Sequence 11, Appl
44	100.5	2.8	1837	3	US-08-928-361B-5	Sequence 5, Appli
45	100.5	2.8	1837	4	US-09-588-995A-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1

US-09-151-189-2  
; Sequence 2, Application US/09151189  
; Patent No. 6232457  
; GENERAL INFORMATION:  
; APPLICANT: Vreeland, Valerie  
; APPLICANT: Ng, Kwan L.  
; APPLICANT: The Regents of the University of California  
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses  
; FILE REFERENCE: 023070-087100US  
; CURRENT APPLICATION NUMBER: US/09/151,189  
; CURRENT FILING DATE: 1998-09-10  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 676  
; TYPE: PRT  
; ORGANISM: Fucus distichus  
US-09-151-189-2

Query Match	100.0%;	Score 3528;	DB 3;	Length 676;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches	676;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0;
QY	1	MLCHAADTTTGGSPMDTGVRLTTSQRAKGRQLEGEKSLGFHPSETPIYKYLEGSET	60	
Db	1	MLCHAADTTTGGSPMDTGVRLTTSQRAKGRQLEGEKSLGFHPSETPIYKYLEGSET	60	
QY	61	WKVKVLPDTGISASKILGKIMARVRIATATAVLAAPCLAFDEVTASGVPEEHKHTGEG	120	
Db	61	WKVKVLPDTGISASKILGKIMARVRIATATAVLAAPCLAFDEVTASGVPEEHKHTGEG	120	
QY	121	RHLQTCNDDALDPTAPNRDNVAFASRRDARRRDGTGTVCQITNGETDLATMFKHS	180	
Db	121	RHLQTCNDDALDPTAPNRDNVAFASRRDARRRDGTGTVCQITNGETDLATMFKHS	180	
QY	181	LPHELGQVTDADFAILEDICILNGDPSICEDVPAGDPAGRLVNPTAAFAIDISGPAFSAT	240	
Db	181	LPHELGQVTDADFAILEDICILNGDPSICEDVPAGDPAGRLVNPTAAFAIDISGPAFSAT	240	
QY	241	TTPPVPTLSPPELAOLAELYNMALARDVPFMOYGYTDEITTTAAANLAGMGFPNLDVAVS	300	
Db	241	TTPPVPTLSPPELAOLAELYNMALARDVPFMOYGYTDEITTTAAANLAGMGFPNLDVAVS	300	
QY	301	IGSDGTVDFFSQLFRATFVGTVETGPFVSQLVNSFTIDAITVBPKEQTFAPDLNMYWDFD	360	
Db	301	IGSDGTVDFFSQLFRATFVGTVETGPFVSQLVNSFTIDAITVBPKEQTFAPDLNMYWDFD	360	
QY	361	EWLNQNGGPPAGPEELDEBELRIFIRNARDIARVSVVDNINTEAYRGLSLLLELGAFSRPG	420	
Db	361	EWLNQNGGPPAGPEELDEBELRIFIRNARDIARVSVVDNINTEAYRGLSLLLELGAFSRPG	420	

Db 361 EWLNIQNGGPPAGPEELDEBLRNFIRNARDLARVSFVDNINTEAYRGSLLILLELGAFSRPG 420  
Qy 421 INGPFFIDSDRQAGFVNFGTSHYFRLIGAELAQRASCYQKWQVHRFARPEALGGTLHNTI 480  
Db 421 INGPFFIDSDRQAGFVNFGTSHYFRLIGAELAQRASCYQKWQVHRFARPEALGGTLHNTI 480  
Qy 481 AGDLDDADFDISLENDLLKRVAEINAAQNPNNVETVLLPQAIQVGSPTHSPSPSGHATQ 540  
Db 481 AGDLDDADFDISLENDLLKRVAEINAAQNPNNVETVLLPQAIQVGSPTHSPSPSGHATQ 540  
Qy 541 NGAFATVVKALIGLDRGCECFPNVPFSDDGLELINFEGACLYEGBEINKLAVNVAFRQ 600  
Db 541 NGAFATVVKALIGLDRGCECFPNVPFSDDGLELINFEGACLYEGBEINKLAVNVAFRQ 600  
Qy 601 MLGIHYRFDGIQGLLGETITVTRLHOELMTFAEATFEPRFLTGEVVKLFQDGTFSIDG 660  
Db 601 MLGIHYRFDGIQGLLGETITVTRLHOELMTFAEATFEPRFLTGEVVKLFQDGTFSIDG 660  
Qy 661 DMCGLVYTGADCOA 676  
Db 661 DMCGLVYTGADCOA 676

RESULT 2  
US-09-596-794-2  
; Sequence 2, Application US/09596794  
; Patent No. 6656715  
; GENERAL INFORMATION:  
; APPLICANT: Vreeland, Valerie  
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases  
; TITLE OF INVENTION: and Their Uses  
; FILE REFERENCE: 023070-087110US  
; CURRENT APPLICATION NUMBER: US/09/596,794  
; PRIOR FILING DATE: 2000-06-19  
; PRIOR APPLICATION NUMBER: US 09/151,189  
; PRIOR FILING DATE: 1998-09-10  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 676  
; TYPE: PRT  
; ORGANISM: Fucus distichus  
US-09-596-794-2

Query Match 100.0%; Score 3528; DB 4; Length 676;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 676; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLCHAADTTGSPMPDTGVLRLTSEORAKGWRQLEGEKSLGFHPSETPYIKYLEGSET 60  
Db 1 MLCHAADTTGSPMPDTGVLRLTSEORAKGWRQLEGEKSLGFHPSETPYIKYLEGSET 60  
Qy 61 WKVKLPDGTGISAKILGKIMARVRIATALAVLAAPCLAFDEVTAGVFPPEBKHTGEG 120  
Db 61 WKVKLPDGTGISAKILGKIMARVRIATALAVLAAPCLAFDEVTAGVFPPEBKHTGEG 120  
Qy 121 RHLQCTNSDDALDPTAPNRDNNVAFASRRDAARRDGTGTCQITNGETDILATMPHKS 180  
Db 121 RHLQCTNSDDALDPTAPNRDNNVAFASRRDAARRDGTGTCQITNGETDILATMPHKS 180  
Qy 181 LPHDELQVNTADDPFAILEDCLINGDFSCEDVPGDPPAGRLVNPTAFAIDISGPAFSAT 240  
Db 181 LPHDELQVNTADDPFAILEDCLINGDFSCEDVPGDPPAGRLVNPTAFAIDISGPAFSAT 240  
Qy 241 TIPPVPTLSSPELAAQLAELYNALARDVPMQYGTDEITTTAAANLAGMGFPNLDVAVS 300  
Db 241 TIPPVPTLSSPELAAQLAELYNALARDVPMQYGTDEITTTAAANLAGMGFPNLDVAVS 300  
Qy 301 IGSDGTVDPSQLFRATFVGVEGTFVSQLLVNSFTTIDAITVEPKQETFPADLNMYMVDFF 360  
Db 301 IGSDGTVDPSQLFRATFVGVEGTFVSQLLVNSFTTIDAITVEPKQETFPADLNMYMVDFF 360

Qy 361 EWLNIQNGGPPAGPEELDEBLRNFIRNARDLARVSFVDNINTEAYRGSLLILLELGAFSRPG 420  
Db 361 EWLNIQNGGPPAGPEELDEBLRNFIRNARDLARVSFVDNINTEAYRGSLLILLELGAFSRPG 420  
Qy 421 INGPFFIDSDRQAGFVNFGTSHYFRLIGAELAQRASCYQKWQVHRFARPEALGGTLHNTI 480  
Db 421 INGPFFIDSDRQAGFVNFGTSHYFRLIGAELAQRASCYQKWQVHRFARPEALGGTLHNTI 480  
Qy 481 AGDLDDADFDISLENDLLKRVAEINAAQNPNNVETVLLPQAIQVGSPTHSPSPSGHATQ 540  
Db 481 AGDLDDADFDISLENDLLKRVAEINAAQNPNNVETVLLPQAIQVGSPTHSPSPSGHATQ 540  
Qy 541 NGAFATVVKALIGLDRGCECFPNVPFSDDGLELINFEGACLYEGBEINKLAVNVAFRQ 600  
Db 541 NGAFATVVKALIGLDRGCECFPNVPFSDDGLELINFEGACLYEGBEINKLAVNVAFRQ 600  
Qy 601 MLGIHYRFDGIQGLLGETITVTRLHOELMTFAEATFEPRFLTGEVVKLFQDGTFSIDG 660  
Db 601 MLGIHYRFDGIQGLLGETITVTRLHOELMTFAEATFEPRFLTGEVVKLFQDGTFSIDG 660  
Qy 661 DMCGLVYTGADCOA 676  
Db 661 DMCGLVYTGADCOA 676

RESULT 3  
US-09-328-352-5503  
; Sequence 5503, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 5503  
; LENGTH: 3892  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-5503

Query Match 3.7%; Score 129.5; DB 4; Length 3892;  
Best Local Similarity 21.0%; Pred. No. 0.0053;  
Matches 158; Conservative 71; Mismatches 245; Indels 279; Gaps 36;

Qy 46 PSETPYIKYLEGSE-----TWKKVLPDGTGISAKILGKIMARVRIAT----- 88  
Db 330 PGSTVTVTVYDGTGSKTVVAGPDGTW---TVPNPGLNDGDEV-----TAVATDPAGNTSG 380  
Qy 89 -ALAVVLA-APCLAFDEV-----TASGV-----FPEBKHTG----- 118  
Db 381 PATAVVDVAVPTVALDVLNDSTPALTGTVNDPTATVTVNVVDVDPYPAVNGDGTWTLA 440  
Qy 119 -----EGRHLQCTNSDDA-----LDPTAPN-----RDNVAFAS 148  
Db 441 DNTLPTLADGPHITITATDAAGNVGDTGCVTVVTDAPNTAGVTTFIDSVTADNVAS 500  
Qy 149 RRDAAARR-----ERDGTGT-VCQITNGETDILATM-----FHKSILPHDELQVNTA 191  
Db 501 --EAGNVTTITGVKLNIPADATNTAVTVWINGVTVYATVNDKTAGTWTVSVEGSL----- 553  
Qy 192 DDFAILDCILNGDFSCEDVPGDPPAGR--LVNPTAFAIDISGPAFSATIPPVPTLS 249  
Db 554 -----VADAKTIDAKVTFDAGNSSTVNDQIYTLDTAAPA--APVIDEV----- 598  
Qy 250 SPELAAQLAELYNALARDVPMQYGTDEITTTAAANLAGMGFPNLDVAVS1--GSDGTV 307  
Db 599 -----NGTDPITGTAEFGSTVTVYTPNGDGTATVAGPDGWSW 634  
Qy 308 D-PFSQLFRATFV-GVETGTFVSQLLVNSFTTIDAITVEPKQETFPADLNMYMVDFF 365

Db 635 SVNPGLNDGDEVEIAITDPAGNPSLPGTATVDA--VGPNTD-----GVNFTVDSVTADNV 688  
Qy 366 QNGGPPAGPEELDELRIRNARDLARVSFVDNINTAYRGSLLILBELGAFSRPGINGPF 425  
Db 689 INASEASGNVTGVLKNV--PADAANTVTVVINGQTYTATV----- 729  
Qy 426 IDSRQAGFVNGTSHYFRLIGAAELAQASCYQKQVHRPARPEALGGTLHNTIAGDLD 485  
Db 730 ---DSTAG-----TWTV-----SVPGSDLT 746  
Qy 486 ADPDISLENDELKRVAEINAAONP---NNEVTYLL---PQA-----I 523  
Db 747 ADAD-----KTIDAKVTFTDAAGNSSVNDTQYITTTAPADPVPINPUNGTDPTGTA 800  
Qy 524 QVGSPTSPYSGHATQNGAFATVLKALIGLDRGCEFPNPFPSDDGLEBLINPEGACLT 583  
Db 801 EPGSTVTVTPDGSTT-----TVVAGPD-GTWTVPNPLNGD----- 837  
Qy 584 YEGEINKLVANVAPGROMLGIHYRFDGIQGLLIGETTITVTLHQLMTPABEATPEPRLP 643  
Db 838 ---KVTIAITDPA--GNPSLPGTATVDVAGPNTDGVNFTVDSVTADNVINASEASGNVTV- 892  
Qy 644 TGEVILFQDGTSGIDGCMGSLVYTVGVADCOA 676  
Db 893 TGVLKNVPADAANTVTVVINGQTYTATVDSTA 925

## RESULT 4

US-09-413-814-67  
; Sequence 67, Application US/09413814  
; Patent No. 6225064  
; GENERAL INFORMATION:  
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
; APPLICANT: Bristol-Myers Squibb, Co.  
; APPLICANT: Bayer, Stefan  
; APPLICANT: Bloecker, Helmut  
; APPLICANT: Brandt, Petra  
; APPLICANT: Cino, Paul M  
; APPLICANT: Dougherty, Brian A  
; APPLICANT: Goldberg, Steven L  
; APPLICANT: Hofle, Gerhard  
; APPLICANT: Mueller, Joachim  
; APPLICANT: Reichenbach, Hans  
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or  
; TITLE OF INVENTION: heteropolyketide compounds  
; FILE REFERENCE: PCT/US 99/23535  
; CURRENT APPLICATION NUMBER: US/09/413,814  
; CURRENT FILING DATE: 1999-10-07  
; EARLIER APPLICATION NUMBER: DE 198 46 493.2  
; EARLIER FILING DATE: 1998-10-09  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 67  
; LENGTH: 2890  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-413-814-67

Query Match 3.5%; Score 123.5; DB 3; Length 2890;  
Best Local Similarity 21.9%; Pred. No. 0.014;  
Matches 109; Conservative 51; Mismatches 185; Indels 153; Gaps 22;  
Qy 158 DGTGTCQITNGETDLMTHFKSLPHDELQGVTDADDFAILDC----- 200  
Db 1989 DGEHFVSSQPLPEPDLAVL-----EAGRVFADLPVLFWCKFAGERLADVLTKTLA 2042  
Qy 201 --IL--NGDFSICEDVPAGDP-----AGRLVNPTAAFAIDISGPAFSATT 241  
Db 2043 LEILFPGSGFDMAERIYRDSPIARYSNGIVGVVESAARVVPAGSMFSILEIGAGTGATT 2102  
Qy 242 IPPVPTLSSPELAQAELAYMALA-----RDVPFMQYGTDEITTTAAANLAGMGGFPN 295  
Db 2103 AAVLPVLLPDRTEYHFTDVSPFLARAEQRFRDYPFLKYGLVDQDPAGQYAHQRFDV 2162

Qy 296 LDAVSGSDGTVPFSQLFRATFVGTVGTFPVSQLVNSFTIDAITYPKQETPAP-DLN 354  
Db 2163 IVAANV-IHATRD-----IRAT-----AKRL-----SLLAPGGL 2192  
Qy 355 YMYD---FDEWLNIQNGPPAGPEELDEELRF-----IRNARD-----LARVSFVDNI----- 399  
Db 2193 VLVEGTGHPWFIDTT-GLIEGWQKYEDDLRIHPLLPARTWCDVLRVGFADAVSLPGD 2251  
Qy 400 -NTEAVRGSLLLELGAFSRPGINGPIDSROAGFVNFSGTSHYFRLIGAAELAQASCY 458  
Db 2252 GSPAGILQGVHLS-----RAGIAGAACDSGGS-----ATESPAARAVR 2292  
Qy 459 QKWQ-----VHRFA-----RPEALGGLTHTNTIAGDL-----DADFDI 490  
Db 2293 QEWADGSADVHRMALERMYFHRPGRQVWVGRLRTGGGAFKALAGDLLLFFDTGOV 2352  
Qy 491 SILENDELLKRVAEINAAQNNEVTYLLPQAQVGSPTSPYSGHAT-----QNG 542  
Db 2353 AEVQGLRLPQLEASAFAPRDPREEWLYALEWQRKODIPEAPAAASSSSAGAWLVLMDOGG 2412  
Qy 543 AFATVLKALIGLDRGGEC 560  
Db 2413 TGAALVSLLEG--RGEAC 2428

## RESULT 5

US-09-335-409-6  
; Sequence 6, Application US/09335409  
; Patent No. 6121029  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/335,409  
; CURRENT FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 3798  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-335-409-6

Query Match 3.5%; Score 123.5; DB 3; Length 3798;  
Best Local Similarity 21.9%; Pred. No. 0.022;  
Matches 109; Conservative 51; Mismatches 185; Indels 153; Gaps 22;  
Qy 158 DGTGTCQITNGETDLMTHFKSLPHDELQGVTDADDFAILDC----- 200  
Db 2760 DGEHFVSSQPLPEPDLAVL-----EAGRVFADLPVLFWCKFAGERLADVLTKTLA 2813  
Qy 201 --IL--NGDFSICEDVPAGDP-----AGRLVNPTAAFAIDISGPAFSATT 241  
Db 2814 LEILFPGSGFDMAERIYRDSPIARYSNGIVGVVESAARVVPAGSMFSILEIGAGTGATT 2873  
Qy 242 IPPVPTLSSPELAQAELAYMALA-----RDVPFMQYGTDEITTTAAANLAGMGGFPN 295  
Db 2874 AAVLPVLLPDRTEYHFTDVSPFLARAEQRFRDYPFLKYGLVDQDPAGQYAHQRFDV 2933  
Qy 296 LDAVSGSDGTVPFSQLFRATFVGTVGTFPVSQLVNSFTIDAITYPKQETPAP-DLN 354  
Db 2934 IVAANV-IHATRD-----IRAT-----AKRL-----SLLAPGGL 2963  
Qy 355 YMYD---FDEWLNIQNGPPAGPEELDEELRF-----IRNARD-----LARVSFVDNI----- 399  
Db 2964 VLVEGTGHPWFIDTT-GLIEGWQKYEDDLRIHPLLPARTWCDVLRVGFADAVSLPGD 3022

```
QY 400 -NTEAYRGSLLLELGAFAFRPGINGPIDSDROAGFVNFGTSHYFRLLIGAAELAQASCY 458
Db 3023 GSPAGILQGHVILS-----RAPGIAGACDSGSES-----ATESPAARAVR 3063
QY 459 QKWQ-----VHRFA-----RPEALGGTLHNTIAGDL-----DADFDI 490
Db 3064 QEWADGSADVHRMALERMVYFHRPGRQVWVHGRLRTGGGAFTKALAGDLLLPEDTQGVV 3123
QY 491 SLENDELLKRVAEINAAQNPNNEVTYLLPQAIQVGSPTHSPYPSGHAT-----QNG 542
Db 3124 AEVQGLRLPQLEASAFAPRDPREWLYALEWQKRDPIPEAPAAASSSSAGAWLVLMDOGG 3183
QY 543 AFATVLKALIGLDRGGEC 560
Db 3184 TGAALVSLLEG--RGEAC 3199

RESULT 6
US-09-568-102-6
; Sequence 6, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 6
; LENGTH: 3798
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-102-6

Query Match 3.5%; Score 123.5; DB 4; Length 3798;
Best Local Similarity 21.9%; Pred. No. 0.022;
Matches 109; Conservative 51; Mismatches 185; Indels 153; Gaps 22;

QY 158 DGTGTCQITNGETDLATMFHKSPLPHDELQGVTTADDFAILEDG-----200
Db 2760 DGEHFVSSQPLPEPDLAVAL-----EAGRVFADLPVLFEWCKFAGERLADVLTKTLA 2813
QY 201 --IL--NGDPSICEDVPAGDP-----AGRLVNPTAAFAIDISGPAFSATT 241
Db 2814 LEILPFGGSFDMARIYRDSPIARYSNGIVRGVVEAARVAVPSGMFSILEIGAGTGATT 2873
QY 242 IPPVPTLSSPELAQAELLYMALA-----RDVPFMOYGTDEITTTAAANLAGMGGFNP 295
Db 2874 AAVLPVLLPDRTEVHTDVSPLFLARAEQRFDPYFLKYGLVDVQBPAGQGVAHQRFDV 2933
QY 296 LDAVSGSDGTVPFSQLFRATFVGTVETGPFVSQLVNSFTIDAITVEPKQETPAP-DLN 354
Db 2934 IVAANV-IHATRD-----IRAT-----AKRL-----SLLAPGGLL 2963
QY 355 YMVD---FDEWLNTONGPPAGPELDEELRF---IRNARD-----LARVSFVDNI-----399
Db 2964 VLVEGTGHPWFDTT-GLIEGWQYEDDLRIDHPLLPARTWCDVLRVGFADAVSLPGD 3022
QY 400 -NTEAYRGSLLLELGAFAFRPGINGPIDSDROAGFVNFGTSHYFRLLIGAAELAQASCY 458
Db 3023 GSPAGILQGHVILS-----RAPGIAGACDSGSES-----ATESPAARAVR 3063
QY 459 QKWQ-----VHRFA-----RPEALGGTLHNTIAGDL-----DADFDI 490
Db 3064 QEWADGSADVHRMALERMVYFHRPGRQVWVHGRLRTGGGAFTKALAGDLLLPEDTQGVV 3123
QY 491 SLENDELLKRVAEINAAQNPNNEVTYLLPQAIQVGSPTHSPYPSGHAT-----QNG 542
Db 3124 AEVQGLRLPQLEASAFAPRDPREWLYALEWQKRDPIPEAPAAASSSSAGAWLVLMDOGG 3183
QY 543 AFATVLKALIGLDRGGEC 560
Db 3184 TGAALVSLLEG--RGEAC 3199
```

```
QY 491 SLENDELLKRVAEINAAQNPNNEVTYLLPQAIQVGSPTHSPYPSGHAT-----QNG 542
Db 3124 AEVQGLRLPQLEASAFAPRDPREWLYALEWQKRDPIPEAPAAASSSSAGAWLVLMDOGG 3183
QY 543 AFATVLKALIGLDRGGEC 560
Db 3184 TGAALVSLLEG--RGEAC 3199

RESULT 7
US-09-567-969-6
; Sequence 6, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 6
; LENGTH: 3798
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-567-969-6

Query Match 3.5%; Score 123.5; DB 4; Length 3798;
Best Local Similarity 21.9%; Pred. No. 0.022;
Matches 109; Conservative 51; Mismatches 185; Indels 153; Gaps 22;

QY 158 DGTGTCQITNGETDLATMFHKSPLPHDELQGVTTADDFAILEDG-----200
Db 2760 DGEHFVSSQPLPEPDLAVAL-----EAGRVFADLPVLFEWCKFAGERLADVLTKTLA 2813
QY 201 --IL--NGDPSICEDVPAGDP-----AGRLVNPTAAFAIDISGPAFSATT 241
Db 2814 LEILPFGGSFDMARIYRDSPIARYSNGIVRGVVEAARVAVPSGMFSILEIGAGTGATT 2873
QY 242 IPPVPTLSSPELAQAELLYMALA-----RDVPFMOYGTDEITTTAAANLAGMGGFNP 295
Db 2874 AAVLPVLLPDRTEVHTDVSPLFLARAEQRFDPYFLKYGLVDVQBPAGQGVAHQRFDV 2933
QY 296 LDAVSGSDGTVPFSQLFRATFVGTVETGPFVSQLVNSFTIDAITVEPKQETPAP-DLN 354
Db 2934 IVAANV-IHATRD-----IRAT-----AKRL-----SLLAPGGLL 2963
QY 355 YMVD---FDEWLNTONGPPAGPELDEELRF---IRNARD-----LARVSFVDNI-----399
Db 2964 VLVEGTGHPWFDTT-GLIEGWQYEDDLRIDHPLLPARTWCDVLRVGFADAVSLPGD 3022
QY 400 -NTEAYRGSLLLELGAFAFRPGINGPIDSDROAGFVNFGTSHYFRLLIGAAELAQASCY 458
Db 3023 GSPAGILQGHVILS-----RAPGIAGACDSGSES-----ATESPAARAVR 3063
QY 459 QKWQ-----VHRFA-----RPEALGGTLHNTIAGDL-----DADFDI 490
Db 3064 QEWADGSADVHRMALERMVYFHRPGRQVWVHGRLRTGGGAFTKALAGDLLLPEDTQGVV 3123
QY 491 SLENDELLKRVAEINAAQNPNNEVTYLLPQAIQVGSPTHSPYPSGHAT-----QNG 542
Db 3124 AEVQGLRLPQLEASAFAPRDPREWLYALEWQKRDPIPEAPAAASSSSAGAWLVLMDOGG 3183
QY 543 AFATVLKALIGLDRGGEC 560
Db 3184 TGAALVSLLEG--RGEAC 3199
```



Db 3184 TGAALVSLLEG--RGEAC 3199

## RESULT 8

US-09-568-480-6  
; Sequence 6, Application US/09568480  
; Patent No. 6355458  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/568,480  
; CURRENT FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 3798  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-568-480-6

Query Match 3.5%; Score 123.5; DB 4; Length 3798;  
Best Local Similarity 21.9%; Pred. No. 0.022;  
Matches 109; Conservative 51; Mismatches 185; Indels 153; Gaps 22;  
Qy 158 DGTGTCQITNGETDLATMFHKSPLHDELQGVTDADDFAILEDG----- 200  
Db 2760 DGEHFVSSQPLPEPDLAVAL-----EAGRVFADLPVLFWECKFAGERLADVLTKTLA 2813  
Qy 201 --IL--NGDFSICEDVPAGDP-----AGRLVNPTAFAIDISGAPFATT 241  
Db 2814 LEILFPGGSFDMARIYRDSPIARYSNGIVRGVESAARVAVPFGMSFISILEIGAGTGATT 2873  
Qy 242 IPPVPTLSSPELAQLAELVWALA-----RDVPFMOYGTDEITTTAAANLAGMGGFPN 295  
Db 2874 AAVLPVLLPDRTEYHFTDVSPLFARAEQRFDRYPFLKYGLLDVDDQEPAGQYAHQRFDV 2933  
Qy 296 LDVAVSIGDGTVPFSQLFRATFVGTVGTPFVSQLLVNSFTIDAIVPEKQETPAP-DLN 354  
Db 2934 IVAANV-IHATRD-----IRAT-----AKRL-----SLLAPGGLL 2963  
Qy 355 YMYD---FDEWLNIONGGPPAGPEELDEELRF---IRNARD---LARVSFVNI----- 399  
Db 2964 VLVEGTGHPFIWFDITT-GLIEGMOKYEDDLRIDHPLLPARTWCDVLRVGFADAVSLPGD 3022  
Qy 400 -NTEAYRGSLLLELGAFSRPGINGPFIQSDROAGFVNFCTSHYFRILIGAAELAQASCY 458  
Db 3023 GSPAGILGQHVILS-----RAGIAGAACDSGSES-----ATESPAARAVR 3063  
Qy 459 QKWQ-----VHRFA-----RPEALGGLTHTNTIAGDL-----DADFDI 490  
Db 3064 QEWADGSADVHRWALERMVYFHRPPGQVWVHGRLRTGGGAFTKALAGDLLLPEDTQGVV 3123  
Qy 491 SLENDELLKRVAEINAAQNPNNEVTVLLPQATQVGSPTSPSPSHAT-----QNG 542  
Db 3124 AEVQGLRLPQLEASAFAPRDPREEWLYALEWQKRDPIPEAPAAASSSSAGAWLVLMQDGG 3183  
Qy 543 AFATVLKALIGLDRGGEC 560  
Db 3184 TGAALVSLLEG--RGEAC 3199

## RESULT 9

US-09-568-486-6  
; Sequence 6, Application US/09568486  
; Patent No. 6355459

; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/568,486  
; CURRENT FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 3798  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-568-486-6

Query Match 3.5%; Score 123.5; DB 4; Length 3798;  
Best Local Similarity 21.9%; Pred. No. 0.022;  
Matches 109; Conservative 51; Mismatches 185; Indels 153; Gaps 22;  
Qy 158 DGTGTCQITNGETDLATMFHKSPLHDELQGVTDADDFAILEDG----- 200  
Db 2760 DGEHFVSSQPLPEPDLAVAL-----EAGRVFADLPVLFWECKFAGERLADVLTKTLA 2813  
Qy 201 --IL--NGDFSICEDVPAGDP-----AGRLVNPTAFAIDISGAPFATT 241  
Db 2814 LEILFPGGSFDMARIYRDSPIARYSNGIVRGVESAARVAVPFGMSFISILEIGAGTGATT 2873  
Qy 242 IPPVPTLSSPELAQLAELVWALA-----RDVPFMOYGTDEITTTAAANLAGMGGFPN 295  
Db 2874 AAVLPVLLPDRTEYHFTDVSPLFARAEQRFDRYPFLKYGLLDVDDQEPAGQYAHQRFDV 2933  
Qy 296 LDVAVSIGDGTVPFSQLFRATFVGTVGTPFVSQLLVNSFTIDAIVPEKQETPAP-DLN 354  
Db 2934 IVAANV-IHATRD-----IRAT-----AKRL-----SLLAPGGLL 2963  
Qy 355 YMYD---FDEWLNIONGGPPAGPEELDEELRF---IRNARD---LARVSFVNI----- 399  
Db 2964 VLVEGTGHPFIWFDITT-GLIEGMOKYEDDLRIDHPLLPARTWCDVLRVGFADAVSLPGD 3022  
Qy 400 -NTEAYRGSLLLELGAFSRPGINGPFIQSDROAGFVNFCTSHYFRILIGAAELAQASCY 458  
Db 3023 GSPAGILGQHVILS-----RAGIAGAACDSGSES-----ATESPAARAVR 3063  
Qy 459 QKWQ-----VHRFA-----RPEALGGLTHTNTIAGDL-----DADFDI 490  
Db 3064 QEWADGSADVHRWALERMVYFHRPPGQVWVHGRLRTGGGAFTKALAGDLLLPEDTQGVV 3123  
Qy 491 SLENDELLKRVAEINAAQNPNNEVTVLLPQATQVGSPTSPSPSHAT-----QNG 542  
Db 3124 AEVQGLRLPQLEASAFAPRDPREEWLYALEWQKRDPIPEAPAAASSSSAGAWLVLMQDGG 3183  
Qy 543 AFATVLKALIGLDRGGEC 560  
Db 3184 TGAALVSLLEG--RGEAC 3199

## RESULT 10

US-09-568-472-6  
; Sequence 6, Application US/09568472  
; Patent No. 6358719  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern

; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/568,472  
; CURRENT FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 3798  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-568-472-6

Query Match 3.5%; Score 123.5; DB 4; Length 3798;  
Best Local Similarity 21.9%; Pred. No. 0.022;  
Matches 109; Conservative 51; Mismatches 185; Indels 153; Gaps 22;  
QY 158 DGTGTVTCQITNGETDLATMFHKSPLPHDELQGVTTADDFAILEDG----- 200  
DB 2760 DGEHFVSSQPLPEPDLAAVL-----EEAGRVFADLPVLFEWCKFAGERLADVLTKGTLA 2813  
QY 201 --IL--NGDFSICEDVPAGDP-----AGRLVNPTAAFAIDISGPAFSATT 241  
DB 2814 LEILPFGSGFDMAERIYRDSPIARYSNGIVRGVVEAARVAVPFGMSFISILEIGAGTGATT 2873  
QY 242 IPPVPTLSSPELAQAELLYMALA-----RDVPFMOYGTDEITTTAAANLAGMGGFNP 295  
DB 2874 AAVLPVLLPDRTEYHFTDVSPLFLARAEORFRDYPFKYGILDVDQEPAGQGYAHORFDV 2933  
QY 296 LDVAVSIGSDGTVPFSQLFRATFVGTVETGPFVSQLVNSFTIDAIVPEPKQETFAP--DLN 354  
DB 2934 IVAANV-IHATRD-----IRAT-----AKRLI-----SLLAPGGLL 2963  
QY 355 YMYD---FDEWLNIQNGPPAGPEELDEELRF---IRNARD-----LARVSFVDNI----- 399  
DB 2964 VLVEGTGHPITWFDITT--GLIEGWQKYEDDLRIDHPLLPARTWCDVLRVGVFADAVSLPGD 3022  
QY 400 -NTEAYRGSILILELGAFAFRPGINGPFIDSDROAGFVNFOTSHVFRILGAELAQRASY 458  
DB 3023 GSPAGILGQHVILS-----RAGIAGACDSSGES-----ATESPAARAVR 3063  
QY 459 QKWQ-----VHRFA-----RPEALGGTLHNTIAGDL-----DADFDI 490  
DB 3064 QEWADGSADVHRMALERMYFHRPGRQVWVHGRLTGGAFTKALAGDLLLPEDTQGVV 3123  
QY 491 SLENDELLKRVAEINAAQNPNNEVTVLLPQAIQVGSPTSPSPSGHAT-----QNG 542  
DB 3124 AEVQGLRLPQLEASAFAPRDPREEWLYALEWQKDPPEAPAAASSSSAGAWLVLMDDQGG 3183  
QY 543 AFATVLKALIGLDRGGEC 560  
DB 3184 TGAALVSLLEG--RGEAC 3199

RESULT 11  
US-09-567-899-6  
; Sequence 6, Application US/09567899  
; Patent No. 6383787  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/567,899  
; CURRENT FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 3798  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-567-899-6

Query Match 3.5%; Score 123.5; DB 4; Length 3798;  
Best Local Similarity 21.9%; Pred. No. 0.022;  
Matches 109; Conservative 51; Mismatches 185; Indels 153; Gaps 22;  
QY 158 DGTGTVTCQITNGETDLATMFHKSPLPHDELQGVTTADDFAILEDG----- 200  
DB 2760 DGEHFVSSQPLPEPDLAAVL-----EEAGRVFADLPVLFEWCKFAGERLADVLTKGTLA 2813  
QY 201 --IL--NGDFSICEDVPAGDP-----AGRLVNPTAAFAIDISGPAFSATT 241  
DB 2814 LEILPFGSGFDMAERIYRDSPIARYSNGIVRGVVEAARVAVPFGMSFISILEIGAGTGATT 2873  
QY 242 IPPVPTLSSPELAQAELLYMALA-----RDVPFMOYGTDEITTTAAANLAGMGGFNP 295  
DB 2874 AAVLPVLLPDRTEYHFTDVSPLFLARAEORFRDYPFKYGILDVDQEPAGQGYAHORFDV 2933  
QY 296 LDVAVSIGSDGTVPFSQLFRATFVGTVETGPFVSQLVNSFTIDAIVPEPKQETFAP--DLN 354  
DB 2934 IVAANV-IHATRD-----IRAT-----AKRLI-----SLLAPGGLL 2963  
QY 355 YMYD---FDEWLNIQNGPPAGPEELDEELRF---IRNARD-----LARVSFVDNI----- 399  
DB 2964 VLVEGTGHPITWFDITT--GLIEGWQKYEDDLRIDHPLLPARTWCDVLRVGVFADAVSLPGD 3022  
QY 400 -NTEAYRGSILILELGAFAFRPGINGPFIDSDROAGFVNFOTSHVFRILGAELAQRASY 458  
DB 3023 GSPAGILGQHVILS-----RAGIAGACDSSGES-----ATESPAARAVR 3063  
QY 459 QKWQ-----VHRFA-----RPEALGGTLHNTIAGDL-----DADFDI 490  
DB 3064 QEWADGSADVHRMALERMYFHRPGRQVWVHGRLTGGAFTKALAGDLLLPEDTQGVV 3123  
QY 491 SLENDELLKRVAEINAAQNPNNEVTVLLPQAIQVGSPTSPSPSGHAT-----QNG 542  
DB 3124 AEVQGLRLPQLEASAFAPRDPREEWLYALEWQKDPPEAPAAASSSSAGAWLVLMDDQGG 3183  
QY 543 AFATVLKALIGLDRGGEC 560  
DB 3184 TGAALVSLLEG--RGEAC 3199

RESULT 12  
US-09-151-189-9  
; Sequence 9, Application US/09151189  
; Patent No. 6232457  
; GENERAL INFORMATION:  
; APPLICANT: Vreeland, Valerie  
; APPLICANT: Ng, Kwan L.  
; APPLICANT: The Regents of the University of California  
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses  
; FILE REFERENCE: 023070-087100US  
; CURRENT APPLICATION NUMBER: US/09/151,189  
; CURRENT FILING DATE: 1998-09-10  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: conserved  
; OTHER INFORMATION: vanadium-binding region 1, amino acids 452-473  
US-09-151-189-9  
Query Match 3.5%; Score 122; DB 3; Length 22;  
Best Local Similarity 100.0%; Pred. No. 3.2e-06;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 452 AQRASYQKQVHRFARPEALG 473

Db 1 AQRASYQKQVHRFARPEALG 22

## RESULT 13

US-09-596-794-9  
; Sequence 9, Application US/09596794  
; Patent No. 6656715  
; GENERAL INFORMATION:  
; APPLICANT: Vreeland, Valerie  
; TITLE OF INVENTION: The Regents of the University of California  
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases  
; TITLE OF INVENTION: and Their Uses  
; FILE REFERENCE: 023070-087110US  
; CURRENT APPLICATION NUMBER: US/09/596,794  
; CURRENT FILING DATE: 2000-06-19  
; PRIOR APPLICATION NUMBER: US 09/151,189  
; PRIOR FILING DATE: 1998-09-10  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 9  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURES:  
; OTHER INFORMATION: Description of Artificial Sequence: conserved Fucus  
; OTHER INFORMATION: vanadium-binding region 1, amino acids 452-473,  
; OTHER INFORMATION: 1st conserved motif  
US-09-596-794-9

Query Match 3.5%; Score 122; DB 4; Length 22;

Best Local Similarity 100.0%; Pred. No. 3.2e-06;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 452 AQRASYQKQVHRFARPEALG 473

Db 1 AQRASYQKQVHRFARPEALG 22

## RESULT 14

US-09-252-991A-32210  
; Sequence 32210, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 32210  
; LENGTH: 1043  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-32210

Query Match 3.3%; Score 117; DB 4; Length 1043;

Best Local Similarity 20.1%; Pred. No. 0.011;

Matches 138; Conservative 83; Mismatches 257; Indels 208; Gaps 34;

QY 52 IKYLEG-----SETWKKVLPDGSISKILGKIMAR-----VRATATAVVL 94

Db 460 VDLDGDCDPVTVGRTYRATNPYPALPDHKILSTKSKYKGRANEIRIDTTAQIS 519

QY 95 AAPCLAFDEVTASGVF-----PEEHKHTGEGRHLQCTCTNSDDALDPTAPNRDNVAF 146

Db 520 AA---LMSDRGASALHGLYTHPRPEGKPRGEGFELRT-----DEHGAVRAAKGLL 568  
QY 147 AS-----RRDAARRRDTGTVCQITNGETDLATMF-----HKSLPHDELGLQVTTADDDFA 195  
Db 569 LSTEEQLRAGAGHLDR---GVVVQVLEAALEARELDGYAGEHQGVGHDAAPQOT----- 620  
QY 196 ILEDICILNGDFSICEDV--PAGDPAGRLVNPTAAFAIDISGPAPSATIIPVPTLSSPEL 253  
Db 621 -LQEA-----RDLGHGANDSGSKSGKGAIA--LSGPAGIAAATPASITLAAGEH 669  
QY 254 AAQLAELYWALARDVPMQYGTDEITTTAAANLAGMGFPNLDVAISGSGTVDPSQL 313  
Db 670 VDSVARQNOQVTAQKVINAGSD-----IGLFAQGG-----EL 703  
QY 314 FRATFVGVEGPFVSOLLVNSFTIDAITVBPKEOTFAPDLNVMYDFDEWLNIQNGGPPAG 373  
Db 704 RQIT-----HOGPMLLQAQKNDIRLEA---EQSEVSASQQHVLVTAKEHTLMCGG----- 752  
QY 374 PEELDEELRFIRNARDLARVSFVDNINTEAY---RGSLLILLELGAFAFRPGINGPFIDSDR 430  
Db 753 -----AYLTLLKGGNI--ELG---MPG----- 768  
QY 431 QAGFVNFGTSHYFRLIGAELAORASCYQKQVHRFARPEALG-----TLHN- 478  
Db 769 --NFVVKAASH--VGAASLEAELEPQEVGETQRREVLKQLDQGTAMPNVPYITTMANG 824  
QY 479 -TTAGDLADDFDISLLENDELLKRVAEINAQNPNEVTYL---LPOAIOVGS--PTHPSY 533  
Db 825 EVIEGVTDAEGATQLQKDMNTAKVDMKTKGPASAVAGIAAAGAAVAVGKLLSGPDA 884  
QY 534 PSGHATQNGAFATVLKALIGLDRGCECFPNPVPFSDGLLELINF--EGACILTYEGEINKLA 592  
Db 885 EAGRALSEGE-----ISLANG-----VFGDSIDYSTVRLRDRDYYVWQCKDYMA 929  
QY 593 VN--VAFGRQMLGI-HYRFDGI--QGLLL-----GEITVTLHQELMTFAEEA 636  
Db 930 PNGHIYFGEELRGVADWSLESQRLQGLFIHEMTHVMOHGHNVLVGVAYQQAQRFLLGD 989  
QY 637 TFEFRLFTGEVVKLFODGTFTSIDGDM 662  
Db 990 QYAYRLEPGTKL---DYNIEQQGDI 1012

## RESULT 15

US-08-474-067-6  
; Sequence 6, Application US/08474067  
; Patent No. 5811518  
; GENERAL INFORMATION:  
; APPLICANT: Ranscht, Barbara  
; TITLE OF INVENTION: T-Cadherin Adhesion Molecule  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: United States  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/08/474,067  
; APPLICATION NUMBER: US/08/474,067  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/213,361  
; FILING DATE: 14-MAY-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/607,293

```
; FILING DATE: 30-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1682
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-474-067-6

Query Match      3.2%; Score 114; DB 2; Length 913;
Best Local Similarity 21.2%; Pred.No. 0.018;
Matches 113; Conservative 65; Mismatches 196; Indels 159; Gaps 28;

Qy 175 TMFHKSLPH---DELQVTDADFAILEDCLINGDFSCIEDVPAGDPAGRLVNP----- 224
Db 409 TVTDKQDPHPAMVARYQMTGGD-----PTGQFTILTDPNNDGLVTVVVPIDFETN 460
Qy 225 ----TAA-----FAIDISGPAFSATTIP-----PVPTLSSPELAAQLAEL 260
Db 461 RMFVLTVAAENQVPLAKGIQHPPQSTATVITVIDVNESPVFVFNPKLVQEGLLAGSM 520
Qy 261 YWMALARDVP-FMQYGTDEITTTAAANLAGMGFPNLDVAVS--IGSDGTVDPPS-----Q 312
Db 521 LTTFTARDPDRYMQQ-----TSLRYSKLSDPANWLKIDPVNGQITTTAVLDRESIYVQNN 575
Qy 313 LFRATFVGVEGTG-PFVVS-----QLLVNSFTIDAITVEPKQ-----ETFAPDLNVMVDFDEW 362
Db 576 MYNATFLASONGIPPMMSGTGTLYLLDINDNAPQVNPKEATTCETLQPNAINITAVDDPD 635
Qy 363 LNIQNGGP-----PAGPEELDEELRIFRNARDLARVS-----FVDNINT 401
Db 636 ID-PNAGPFAPFELPDSPPSIKRNWTIVIRISGDHAQLSLRIFLEAGIYDVPIVITDSGNP 694
Qy 402 EAYRGSLLLELGFARPGINGPIDSDRQAGFVNFTSHYFR-LIGAAELAQKASCYQK 460
Db 695 HASSTSVLKVKV---CQCDINGDCTDVRIVG-AGLGTGAIITAILLCIIILLILVLMFVV 750
Qy 461 WQVHRFARPEAL-----GGTLHTIAGDLADDFDISLLENDELLK 500
Db 751 WMKRDKERQAKQLLIDPDDVRNLIKYEDEGG-----GEEDQYDLSQLQOPTVE 803
Qy 501 RVAEINAAQPNNEVTYLLPQAIQ-VG-----SPTH--PSYP-----SGHATQNGAFATV 547
Db 804 -----PDAIKPVGIRRLDERPIHAEQYPVRSAAHPHGDIGDFINE 844
Qy 548 LKALIGLDRGCEFPNVPFSDGDLLEINFEACLYTEGINKLVNVAFRQ 600
Db 845 -----GLAKAAD--NDPTAPPYDSLIVFDYEGSGST-AGSLSSLNSSSGGEQ 889
```

Search completed: September 19, 2004, 04:00:32  
Job time : 31.3913 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 19, 2004, 04:03:44 ; Search time 1014 Seconds  
(without alignments)  
3364.964 Million cell updates/sec

Title: US-10-691-383-2  
Perfect score: 3528  
Sequence: 1 MLCHAADTRGSPMPDTGV.....SIDGDMCSGLVYTGVAQCQA 676

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3327077 seqs, 2523723180 residues

Total number of hits satisfying chosen parameters: 6654154

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US10691383/runat.17092004.102657.1795/app\_query.fasta\_1.1230  
-DB=Published Applications NA -QFMT=fastap -SURFIX=rnnpb -MINWATCH=0.1  
-LOOPCUL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsu62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-MAXLEN=2000000000 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-NCPU=6 -ICPU=3 -NO MWAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US09\_PUB.seq2:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

RESULT 1  
US-09-840-762A-1  
; Sequence 1, Appli US-09-840-762A-1  
; Patent No. US20020035245A1  
; GENERAL INFORMATION:  
; APPLICANT: Vreeland, Valerie  
; APPLICANT: Ng, Kwan L.  
; APPLICANT: The Regents of the University of California  
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses  
; FILE REFERENCE: 023070-08710005  
; CURRENT APPLICATION NUMBER: US/09/840, 762A  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR FILING DATE: 09/151,189  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 2931  
; TYPE: DNA  
; ORGANISM: Fucus distichus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (228)..(2258)

ALIGNMENTS

1	3528	100.0	2931	9	US-09-840-762A-1	Sequence 1, Appli
2	3528	100.0	2931	17	US-10-691-383-1	Sequence 1, Appli
3	131	3.7	9579	15	US-10-132-134-11	Sequence 11, Appli
4	131	3.7	52101	15	US-10-132-134-1	Sequence 11, Appli
5	129	3.7	1236	17	US-10-437-963-29553	Sequence 29553, A
6	128.5	3.6	1953	16	US-10-369-493-42356	Sequence 42356, A
7	124.5	3.5	7185	15	US-10-329-079-48	Sequence 48, Appli
8	124.5	3.5	61944	13	US-10-329-079-34	Sequence 34, Appli
9	124	3.5	18930	13	US-10-282-122A-31609	Sequence 31609, A
10	123.5	3.5	68750	14	US-10-014-717-1	Sequence 1, Appli
11	123	3.5	5820	13	US-10-152-886-44	Sequence 44, Appli
12	122	3.5	2843	13	US-10-194-163-21	Sequence 21, Appli
13	122	3.5	3969	15	US-10-156-761-2515	Sequence 2515, Ap
14	122	3.5	9025608	15	US-10-156-761-1	Sequence 1, Appli
15	121	3.4	5040	17	US-10-437-963-65478	Sequence 65478, A
16	120	3.4	2664	15	US-10-156-761-1241	Sequence 1241, Ap
17	118	3.3	2256846	17	US-10-470-585-1	Sequence 1, Appli
18	117.5	3.3	3505	13	US-10-282-122A-23847	Sequence 23847, A
19	117	3.3	53500	10	US-09-953-348-76	Sequence 76, Appli
20	117	3.3	53500	15	US-10-267-255-76	Sequence 1, Appli
21	116.5	3.3	1830121	15	US-10-329-960-1	Sequence 1, Appli
22	116.5	3.3	1830121	16	US-10-329-670-1	Sequence 1, Appli
23	116.5	3.3	9025608	15	US-10-156-761-1	Sequence 1, Appli
24	116	3.3	5352	17	US-10-437-963-29803	Sequence 29803, A
25	116	3.3	15738	15	US-10-329-079-12	Sequence 12, Appli
26	116	3.3	37360	15	US-10-329-079-6	Sequence 6, Appli
27	115	3.3	2193	17	US-10-437-963-19583	Sequence 19583, A
28	115	3.3	64492	13	US-10-378-083-1	Sequence 1, Appli
29	115	3.3	125401	17	US-10-203-293-35	Sequence 35, Appli
30	114	3.2	972	17	US-10-317-271A-12	Sequence 12, Appli
31	114	3.2	7095	15	US-10-156-761-7340	Sequence 7340, Ap
32	114	3.2	11007	16	US-10-402-842-5	Sequence 5, Appli
33	114	3.2	11402	17	US-10-672-396-10	Sequence 10, Appli
34	114	3.2	47988	16	US-10-402-842-1	Sequence 1, Appli
35	114	3.2	2256846	17	US-10-470-585-1	Sequence 1, Appli
36	113.5	3.2	2586	16	US-10-369-493-31958	Sequence 31958, A
37	113.5	3.2	2824	9	US-09-905-983-1	Sequence 1, Appli
38	113.5	3.2	2824	9	US-09-905-983-3	Sequence 3, Appli
39	113.5	3.2	10419	14	US-10-027-806-3	Sequence 3, Appli
40	113.5	3.2	10419	14	US-10-034-623-3	Sequence 3, Appli
41	113.5	3.2	10419	15	US-10-027-801-3	Sequence 3, Appli
42	113.5	3.2	10419	15	US-10-029-120-3	Sequence 3, Appli
43	113.5	3.2	42432	14	US-10-027-806-2	Sequence 2, Appli
44	113.5	3.2	42432	14	US-10-034-623-2	Sequence 2, Appli
45	113.5	3.2	42432	15	US-10-027-801-2	Sequence 2, Appli

OTHER INFORMATION: vanadium bromoperoxidase  
US-09-840-762A-1

## Alignment Scores:

Pred. No.: 0 Length: 2931  
Score: 3528.00 Matches: 676  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

US-10-691-383-2 (1-676) x US-09-840-762A-1 (1-2931)

```
QY 1 MetLeuCysHisAlaAlaAspThrThrArgGlySerProMetProAspThrGlyValLeu 20
DB 228 ATGCTTTGGCATGCGCGGACACGACAAAGAGGCTCTCTATGCTGACACCGAGTGCTT 287
QY 21 ArgLeuLeuThrSerGluGlnArgAlaLysGlyTTPArgArgGlnLeuGluGlyLys 40
DB 288 CGGTTGCTCATATCAGACGAGCGGCTTAAGGTTGGAGACGCCAGTTAGAGGGGAGAAA 347
QY 41 SerLeuGlyPheHisProSerGluThrProTyrIleLysTyrLeuGluGlySerGluThr 60
DB 348 TCACCTAGGTTTTCATCAACGACGACGCTTATATCAAGTACTTGAAGGCTCTGAGACT 407
QY 61 TrpLysLysValLysLeuProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIle 80
DB 408 TGAAGAAGGTTAAGCTTCCACGAGCGGATATCGGCTTCCAGATCTCGGGTAAAAAT 467
QY 81 MetAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAla 100
DB 468 ATGGCCAGGTCGCGATCGTACCGCTTGGCTGTGTAATGCGCCGACCCCTGTTGGCA 527
QY 101 PheAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGlyGluGly 120
DB 528 TTCGACGAGTCAACGACGAGTGTGTTTCCCTCAGGAAACACAAAGCACACCGGGGGGA 587
QY 121 ArgHisLeuGlnThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArg 140
DB 588 AGACACCTCCAGACCTGTGCAAACTCCGACGATGCGCTGGATCCGACGGCGCGCAATAGA 647
QY 141 ArgAspAsnValAlaPheAlaSerArgArgAlaAlaAlaArgArgGluArgAspGlyThr 160
DB 648 AGGACACACCTAGCTTTTGGTTCGCGCGCGATGCCGCCAGGCGAGAACCTGACGGGACA 707
QY 161 GlyThrValCysGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSer 180
DB 708 GGGACTGCTCTGCCAAATCACTAACGGAGAACTGATTTGGCTACCATGTTCCCAAGTCT 767
QY 181 LeuProHisAspGluLeuGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCys 200
DB 768 CTGCCACCATGAACATGGGACAGGTAAACCGACAGCTTCGCTATCTCTGAGGACTGC 827
QY 201 IleLeuAsnGlyAspPheSerIleCysGluAspValProAlaGlyAspProAlaGlyArg 220
DB 828 ATCTTAAACCGAGATTTTCAGCATTTGCGAGACGTGCTCGCGGAGACCCGCGGGTCCG 887
QY 221 LeuValAsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThr 240
DB 888 CTGCTCAATCTCACTCGCTGGGTTGCCATCATATATCCGGTCCCGCATTTCTGGGTACG 947
QY 241 ThrIleProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeu 260
DB 948 ACAATACCCCGGTACCTACCTTTCTCTGAGCTCCGCGCTCAGTTGGCGGAGCTA 1007
QY 261 TyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGluIleThr 280
DB 1008 TACTGGATGGCGTGGCCGAGATGTACCTTTATGTCAGTATGGCACCGACGAAATTACC 1067
QY 281 ThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSer 300
DB 1068 ACTACCGCGGACCACTCGCTGGAAATGGGAGGCTTCCCAAATCTGGACGCGCTGTCG 1127
```

```
QY 301 IleGlySerAspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGly 320
DB 1128 ATAGGGTCCGATGGTACGGTGGACCCGTTCTCCAGCTCTTCGAGGCGACTTCGTGGT 1187
QY 321 ValGluThrGlyProPheValSerGlnLeuValAsnSerPheThrIleAspAlaIle 340
DB 1188 GTTGAACCGGGGCCCTTTGTCTCTCAGCTGCTCGTGAACAGCTTCACCATCGACGTATT 1247
QY 341 ThrValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPheAsp 360
DB 1248 ACGGTGCAACCGAAGCAGGAGACATTCGCCCGGACTTGAACATATATGGTCGATTTGAC 1307
QY 361 GluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluGlu 380
DB 1308 GAATGGCTGAACATTCAGATGTGTGACCCCGCGCGGCCCGGAGATTAGACGAAGAG 1367
QY 381 LeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsn 400
DB 1368 CTGCGTTTATCGTAACGCCCGGACCTGCGGAGGCTCTCTTCGTGGACAATATCAAC 1427
QY 401 ThrGluAlaTyrArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGly 420
DB 1428 ACCGAAGCTTATCGCGGCTCTTATCTACTTTCAGCTGGAGGCTTCGACGAGGCCGCT 1487
QY 421 IleAsnGlyProPheIleAspSerAspArgGlnAlaGlyPheValAsnPheGlyThrSer 440
DB 1488 ATCAACGGTCCATTCATCGACAGTGATCGCAGCGCGGCTTCGTCAACTTCGCGACGCT 1547
QY 441 HisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLys 460
DB 1548 CACTACTTCAGATTAGTGGTCCGCGAGCTGCGCGAGCTGCTCGTTACTTACCAAAAG 1607
QY 461 TrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIle 480
DB 1608 TGCAGCGTGCATCGATTGTCACCGCCCGAGGCTCTCGGGGGTACCTTCCACCAACCATC 1667
QY 481 AlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLys 500
DB 1668 GCGGGGATCTAGATGCGACATTCGACATCTCCCTTCTTGAAATGATGAGCTCTTGAAA 1727
QY 501 ArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrThrLeuLeuPro 520
DB 1728 CGTGTGCGGAGATAAATGCGCGCAGAAATCCCAACAAACAGAGTCACTTCTTCTTCCA 1787
QY 521 GlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGln 540
DB 1788 CAAGCTATCCAACTGGGATCGCAACGACCTTCTTACCCGCGCGGACGCTTACCCAA 1847
QY 541 AsnGlyValaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyCys 560
DB 1848 AATGGAGCATTTGCCACAGTTCTGAAGGCCCTCATTTGGCCCTAGATCGGGAGGTGAGTGC 1907
QY 561 PheProAsnProValPheProSerAspGlyLeuGluLeuIleAsnPheGluGlyAla 580
DB 1908 TTCCTTAAACCCCGTGTTCACAGCGATGACGGCTGGAACTAATACTCACTTCGAGGGGCA 1967
QY 581 CysLeuThrThrGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGln 600
DB 1968 TGCTTACATATGAGGAGAGATCAACAGCTCGGCTCAACGTCGATTTGGAGGCGAG 2027
QY 601 MetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThrIle 620
DB 2028 ATGCTGGGCATCACTATCGTTCGACGGTATCCNAGGCCCTACTTCTCGGAGAGACAATC 2087
QY 621 ThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluAlaThrPheGluPhe 640
DB 2088 ACTGTACGAACACTTCACAGGAGCTGATACGTTCCGCCGAGGAAGCCACTTTGAATTC 2147
QY 641 ArgLeuPheThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIleAspGly 660
DB 2148 CGCTTATTCACCGGAGGTCATCAAACTTTTCAGGACGGGACATTTCTCATCGATGGA 2207
QY 661 AspMetCysSerGlyLeuValTyrThrGlyValAlaAspCysGlnAla 676
```

```
Db      2208  GATATGTTCCGGTTTGGTTTACACTGGCGTGGCAGCTGCCAGGCT 2255
|||||
US-10-691-383-1
; Sequence 1, Application US/10691383
; Publication No. US20040110260A1
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; TITLE OF INVENTION: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 023070-087110US
; CURRENT APPLICATION NUMBER: US/10/691,383
; CURRENT FILING DATE: 2003-10-21
; PRIOR FILING DATE: US/09/596,794
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: US 09/151,189
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2931
; TYPE: DNA
; ORGANISM: Fucus distichus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (228)..(2258)
; OTHER INFORMATION: vanadium bromoperoxidase
US-10-691-383-1
Alignment Scores:
Pred. No.: 0 Length: 2931
Score: 3528.00 Matches: 676
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 17 Gaps: 0

US-10-691-383-2 (1-676) x US-10-691-383-1 (1-2931)
Qy      1  MetLeuCysHisAlaAlaAspThrArgGlySerProMetProAspThrGlyValLeu 20
Db      228  ATGCTTTGCCATGAGCGGACACACAAAGAGCTCTCCTATGCTGACACCGGAGTGT 287
Qy      21  ArgLeuLeuThrSerGluGlnArgAlaLysGlyTrpArgArgGlnLeuGluGlyLys 40
Db      288  CGGTGCTCATCATCAGACGACGCGCTAAAGTTGGAGACCCAGTTAGAGGGGGAGAA 347
Qy      41  SerLeuGlyPheHisProSerGluThrProTyrIleLysTyrLeuGluGlySerGluThr 60
Db      348  TCACTAGTTTTCATCCAAAGCGAGACGCTTATATCAAGTACTTGGAGGCTCTGAGACT 407
Qy      61  TrpLysLysValLysLeuProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIle 80
Db      408  TGGAAAGAGTTAAGCTTCCAAACGACGCGCATATCGGCTTCCAAGATCCTGGGTAAAT 467
Qy      81  MetAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAla 100
Db      468  ATGCCACAGGTCCTGACATCGCTTACCGCTTGGCTGTGGTACTGGCGGACACCTGTGGCA 527
Qy     101  PheAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGlyGluGly 120
Db     528  TTCACGAGGTTCACAGCCAGTGGTGTTCCTCGAGGACACACAGACACCGGGAGGGA 587
Qy     121  ArgHisLeuGlnThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArg 140
Db     588  AGACACCTCCAGACCTGTACAAATCCGACGATCGCTGGATCCGACGCGGCCCAATAGA 647
Qy     141  ArgAspAsnValAlaPheAlaSerArgArgAspAlaAlaArgArgGluArgAspGlyThr 160
Db     648  AGGACAACTAGTCTTTGCGTCGCGCGCGATGCCGCCAGCGAGGAGAACGTCGCGGACA 707

161  GlyThrValCysGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSer 180
708  GGGACTGTCTGCCAAATCACTAACGGAGAACTGATTTGGCTAGCATGTTCCACAAGTCT 767
Qy     181  LeuProHisAspGluLeuGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCys 200
768  CTGCCACAGATGAACCTGGGACAGTAACCGACAGCACTTCGGCTATCTCGAGGACTGC 827
Qy     201  IleLeuAsnGlyAspPheSerIleCysGluAspValProAlaGlyAspProAlaGlyArg 220
828  ATCTTAAACGGAGATTTTCAGCATTTGCGAGACGCTGCTCGGGAGACCCGCGGCTCGC 887
Qy     221  LeuValAsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThr 240
888  CTCGTCAATCTACCGCTGGCTTTGCCATCAGCATATCCGCTCGCGCATTTCTCGGCTACG 947
Qy     241  ThrIleProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeu 260
948  ACAATACCCCGGTACCTTCCCTTCTCCTGAGCTCGCGCTCAGTTGGCGGAGCTA 1007
Qy     261  TyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGluIleThr 280
1008  TACTGGATGGCGCTGGCCAGGATGATACCTTTATGAGATGGCACCGACGAAATTACC 1067
Qy     281  ThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSer 300
1068  ACTACCGCGGCGACCAACCTCGCTGGAATGGAGGCTTCCCAATCTGGACGCGCTGTCG 1127
Qy     301  IleGlySerAspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGly 320
1128  ATAGGGTCCGATGGTACGGTGGACCCGTTCTCCAGCTTCTCCGAGCGACCTTCGTTGGT 1187
Qy     321  ValGluThrGlyProPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIle 340
1188  GTTGAACCGGGCCCTTTGCTCTCAGCTGCTGCTGAACAGCTTCACCATCGACGCTATT 1247
Qy     341  ThrValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPheAsp 360
1248  ACGGTCGAACCGAAGCAGGAGACATTCGCCGCCGACTTGAACCTATATGCTCGATTTTGC 1307
Qy     361  GluTrpLeuAsnIleGlnAsnGlyGlyProAlaGlyProGluGluLeuAspGluLeu 380
1308  GAATGGCTGAACATTCAGAATGGTGACCCCGCGCGCGCTTGAAGAGTTAGACGAAGAG 1367
Qy     381  LeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsn 400
1368  CTGGCTTTATCCGTAAACCGCGCGGACCTGGCAGGGTCTCTCTCGTGGACAATATCAAC 1427
Qy     401  ThrGluAlaTyrArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGly 420
1428  ACCGAAGCTTATCGCGGCTCTCTTATCTACTTGGCTGGAGGCTTCAGCAGGCGCGCT 1487
Qy     421  IleAsnGlyProPheIleAspSerAspArgGlnAlaGlyPheValAsnPheGlyThrSer 440
1488  ATCAACCGCTCATTCATCGACAGTATCGGAGCGCGGCTTCGTCACCTTCGGCACGCTCT 1547
Qy     441  HisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLys 460
1548  CACTACTTCAGATTGATAGGTGCGCGCGAGCTGGCGAGCGTGGCTCGTGTATCCAAAG 1607
Qy     461  TrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIle 480
1608  TGGCAGGTGCTATGATTTGACGCGCGCGAGGCTCTCGGGGGTACCCCTCCACAACACCATC 1667
Qy     481  AlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLys 500
1668  GCGGGGATCTAGATGCGAGCTTCGACATCTCCCTCTTTGAAATGATGAGCTCTTGAA 1727
Qy     501  ArgValAlaGluIleAsnAlaAlaGlnProAsnAsnGluValThrTyrLeuLeuPro 520
1728  CGTGTGGCGGAGTAAATGGCGCGAGATCCCAACACAGAGGTCACCTACTCTTCTTCCA 1787
Qy     521  GlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGln 540
```

```
Db 1788 CAGCTATCCAGTGGGATCGCCACACGACCTTCTACCGTCCGGCGACGCTACCCAA 1847
Qy 541 AsnGlyAlaPheAlaThrValLeuLysAlaLeuLeGlyLeuAspArgGlyGlyCys 560
Db 1848 AATGAGCAATTGCGACAGTTCGAGGCCCTCATTTGGCCTAGATCGGGAGGTGAGTGC 1907
Qy 561 PheProAsnProValPheProSerAspAspGlyLeuGluLeuLeuLeuLeuLeuLeu 580
Db 1908 TTCCCTTAACCCCGGTTCCTCAAGCGATGAGCGCTGGAATCAATCACTTCGAAGGGCA 1967
Qy 581 CysLeuThrTyrgluGlyGluLeuAsnLysLeuAlaValAsnValAlaPheGlyArgGln 600
Db 1968 TGCTTACATATGAGGAGAGATCAACAGCTCGCGTCAACGTGCGATTTGGAGGCGAG 2027
Qy 601 MetLeuGlyLeuHisTyrgPheAspGlyLeuGlnGlyLeuLeuLeuGlyGluThrIle 620
Db 2028 ATGCTGGGCATCCACTATCGGTTCGACGGTATCCAGGCTACTTCTCGGAGACAAATC 2087
Qy 621 ThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluAlaThrPheGluPhe 640
Db 2088 ACTGTAGCAACATTCACCGAGGAGCTGATGACGTTCGCCGAGGAAGCCACTTTGAATC 2147
Qy 641 ArgLeuPheThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIleAspGly 660
Db 2148 CGCTTATTCACCGAGAGGTCAACAACCTTTCCAGACGGGACATTTCTCCATCGATGA 2207
Qy 661 AspMetCysSerGlyLeuValTyThrGlyValAlaAspCysGlnAla 676
Db 2208 GATATGTGTTCCGGTTTGGTTTACACTGGCGTGGCGACTGCCAGGCT 2255
```

## RESULT 3

US-10-132-134-11

; Sequence 11, Application US/10132134

; Publication No. US20030171562A1

; GENERAL INFORMATION:

; APPLICANT: Farnet, Chris

; APPLICANT: Yang, Xianshu

; APPLICANT: Staffa, Alfredo

; APPLICANT: Zazopoulos, Emmanuel

; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES

; FILE REFERENCE: 3012-2US

; CURRENT APPLICATION NUMBER: US/10/132,134

; CURRENT FILING DATE: 2002-04-26

; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: Patentin version 3.0

; SEQ ID NO 11

; LENGTH: 9579

; TYPE: DNA

; ORGANISM: Streptomyces platensis subsp. roseaceus

US-10-132-134-11

```
Alignment Scores:
Pred. No.: 0.000459 Length: 9579
Score: 131.00 Matches: 152
Percent Similarity: 30.12% Conservative: 82
Best Local Similarity: 19.56% Mismatches: 254
Query Match: 3.71% Indels: 289
DB: 15 Gaps: 38
```

US-10-691-383-2 (1-676) x US-10-132-134-11 (1-9579)

```
Qy 5 AlaAlaAspThrThrArgGly---SerProMetPro-AspThrGlyValLeuArgLeuLe 23
Db 6726 GCCGACATGACGACGCCGCCCGCGACCTTCGACCCGGGACGACTTTCACCCCGCTCGG 6785
Qy 23 uThrSerGluGlnArgAlaLysGlyTyArgArgGlnLeuGluGlyGlyLysSerLeuG1 43
Db 6786 CACCGCGGGCTCGACGCCCGCGCGTACGCGGCGAGCGTGCAGAGGGTTCACCGACTCGA 6845
Qy 43 yPheHis-----ProSer-----GluThrProTyrlleLy 53
Db 6846 CGCCCAACCGGTGCTGTCGTACGGGTTCGCGAACCAGGCGCATGTGCCAGGACCCGACAGCA 6905
```

```
Qy 53 sTyrlLeuGluGlySerGluThrTrpLysLysValLysLeuProThrAspGlyIleSer-- 72
Db 6906 GCACGTGCAT---CTCCCGCCCTGGCGCTCGTCGGGCTG---ACCAGGGTGTTCAGCA 6959
Qy 73 -----AlaSerLysIleLeuGlyLysIleMe 81
Db 6960 CGCGTGGGGCGGGCCGACGCCCGCTGGTGGTCCGATCCGTGCGAGGCGGAGCAGTG 7019
Qy 81 tAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPh 101
Db 7020 GGAGCGCACCCCG---GCGATCGTCTGCGCGG----- 7050
Qy 101 eAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGly----- 118
Db 7051 -----ACGTCCGACGCCGCTTC-----CATCGGCTTTCCTCGA 7085
Qy 119 ----GluGlyArgHisLeuGlnThrCysThrAsnSerAsp-----AspAlaLeuAs 134
Db 7086 CGAGGACGGCGCGTCTGGCGCGGGTCCGAGGACGCCGAGTTCACCGCGGCGGACCTTGA 7145
Qy 134 pProThrAlaProAsnArgArgAspAsnValAlaPheAla----- 147
Db 7146 GCGGCACTCCCGGTGAGCGCGACGCGCACTCTGTCGCACTGCCGAGCGCTCGCGTCC 7205
Qy 148 -----SerArgArgAspAlaAlaArgArgG1 156
Db 7206 GGTGCTGGAGACGCCGCTGTACGGGGGAGTGCGACGACGTGCGAGGCCGTGCGCGCGA 7265
Qy 156 uArgAspGlyThrGlyThrValCysGlnIleThrAsnGly----- 169
Db 7266 GCGGAGCGCTCGTGACCGTTCGGCGGCTCGCGAGCGGCGCGGCGGCTCGTCGCGTC 7325
Qy 170 -----GluThrAspLeuAlaTh 175
Db 7326 GCTGCGGACGCGTCCGACCTGCTCAAGTTTCGACCTGCGCGACATCGACCTCGACAC 7385
Qy 175 rMetPheHisLys---SerLeuProHisAspGluLeuGlyGlnValThrAlaAspAspPh 194
Db 7386 GCATCTCCACGCGTACGCGTTCGAGTCCATCGCGCTGGCCAACTGCGCTCGGACTCAA 7445
Qy 194 eAlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAspValProAl 214
Db 7446 CGCGCTCTCGGACGACCTCACCCCGCGCTCTTCGAGTGTCTCGAC----- 7497
Qy 214 aGlyAspProAlaGlyArgLeuValAsnProThrAlaAlaPheAlaIleAspIleSerG1 234
Db 7498 -----ATCCGACGCTCGCGAGTACCTGCTCGACCGCTACGG 7535
Qy 234 yProAlaPheSerAlaThrThr-----IleProProValProThrLeuSe 249
Db 7536 CCCGAGCTGAGCTCTCCACGAGCGCGAGCGGCCCGCGCTCGCGCCGCCCGGCC 7595
Qy 249 rSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyTrpMetAlaLeuAlaArgAspVa 269
Db 7596 GTCCCA-----GT 7604
Qy 269 lProPheMetGlnTyrglyThrAspGluIleThrThrAlaAlaAlaAsnLeuAlaG1 289
Db 7605 GCGATGCGCGCACCCCGCGCGGCGGACGAC-----GACGCGGTGCGCATCGTCGCGCG 7655
Qy 289 yMetGlyGlyPhePheAsnLeuAspAlaValSerIleGlySerAspGlyThrValAspPr 309
Db 7656 TGGCGAGCGTTCCCGCGCGGAC-----GACCTGGGACAC 7691
Qy 309 oPheSerGlnLeuPheArg-----AlaThrPheValGly-----Va 321
Db 7692 CTTCGCGACAGCTGCGCGCGGCGGAGGACCTGATCGCGACTACCCCGGCGACCGCTT 7751
Qy 321 lGluThrGlyProPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIleTh 341
Db 7752 CGACGGGGGCGCTACGCGGAGGTGCTGCGCGCGGCGGAGCTTCCCGAAGTTTCCGCGCG 7811
```



```
Qy 341 rValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPheAspG1 361
Db 7812 GATCGAG---GGCGTGGACCGCTTCGACCGCGACTTCTTCACCTGTCGGCTGGAGGC 7868
Qy 361 u-----TrpLeuAsnIleGlnAs 367
Db 7869 GGAGCTGATGACCGGAGACACCGGCTGGGCTGGAGACCGTGTGGCGCGCTGGAGAA 7928
Qy 367 nGlyGly-----ProPheAlaGlyProGluGluLeu-----As 378
Db 7929 CGGCGGTACCGCCCGCGCGCTCCCGAGACACCGGCTTACTTCGGCGTCTCCGG 7988
Qy 378 pGluGluLeuArgPheIleArgAsnAlaArgAspLeuAla----- 391
Db 7989 CAGCGACTACCAACCTGCTCAACGCGCAGTGGCGTGGCACCAGCGCTTTCACCGGCAC 8048
Qy 392 -----ArgValSerPheValAspAsnIleAsn----- 400
Db 8049 CGGCAACGCCACTCGATGTGGCAACCGGATCTCTTACGTCTGGACGTGACCGGCC 8108
Qy 401 -----ThrG1 402
Db 8109 GAGCGAACCCTGCACACGGCTGCTCCAGTCTGCTCGCGCTGCACCGCGCTCGA 8168
Qy 402 uAlaTyArg-----GlySerLeuIleLeuLeuG1 412
Db 8169 GCACATCCGTCGCGCGCGATCGAGATGGCCATCGCGCGCTGTCAACCTGCTGTGAG 8228
Qy 412 uLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSerArgGlnAl 432
Db 8229 CGTGAGACACCTTCGCCGCGAGCACATGCGCGGC---ATGCTCAGCCCGACGCGCGCTG 8285
Qy 432 aGlyPheValAsnPheGlyThrSerHisTyrPheArgLeuIleGly---AlaAlaGluLe 451
Db 8286 CAAGACCTTCTCCCGCGCGGAGCGGTACGTCTCGTCCGAGGCGCTCGCGCGGTCT 8345
Qy 451 uAlaGlnArgAlaSerCysTyrGlnLys-----TrpGlnValHisArgPh 466
Db 8346 GCTCAAGCGCTCGCCCGGCGAGCGGACGCGGACGCCATCTGGGCGCTGCTCCGGG 8405
Qy 466 eAlaArg----- 468
Db 8406 CAGCGCGAGAACCAACGCGCGCGCGCGGTTCGCTGACCGCCCGAACGCGAAGCGCA 8465
Qy 469 -----ProGluAlaLeuGlyGlyThr-- 475
Db 8466 GGCGCGCTGATCCAGACGCCATCGCGCGCATCGACCGGACGATCGGCTACGTCGA 8525
Qy 476 -LeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuGluAs 495
Db 8526 GGCGCAGCGCACCGGCAACCGC---CTGGCGACCGCGTCAAGGTCAACGCCCTCGACAG 8582
Qy 495 nAspGluLeuLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluVa 515
Db 8583 CGCTACCGCGCTGCGCACCGCGAG----- 8610
Qy 515 lThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSe 535
Db 8611 -----GGCGCGCGCGCGCACCGCGCGCGCGCGCGCGGTG 8639
Qy 535 r-----GlyHisAlaThrGlnAsnGlyAlaPheAlaTh 546
Db 8640 CGCGCTCGGCTCGGTGAAGACCAACATCGGCAACCGGAGTCGGCGCGCGCGCTGGCGG 8699
Qy 546 rValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCysPhePro----- 562
Db 8700 AGTCTGAAGGTGCTGCTCGCCATG---CGTACCGCGGAGTGC CGCGCGCTTGCATG 8756
Qy 563 -----AsnProValPheProSerAspAspGlyLeuLeuIle 575
Db 8757 CGACCGGCTCAACCGGACCTGCGCTCGACCGCGGATTCGAGGTGTA 8805
```

RESULT 4

```
US-10-132-134-1
; Sequence 1, Application US/10132134
; Publication No. US20030171562A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Staiffa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES
; FILE REFERENCE: 3012-20S
; CURRENT APPLICATION NUMBER: US/10/132,134
; CURRENT FILING DATE: 2002-04-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 52101
; TYPE: DNA
; ORGANISM: Streptomyces platensis subsp. roseaceus
US-10-132-134-1
```

```
Alignment Scores:
Pred. No.: 0.00725 Length: 52101
Score: 131.00 Matches: 152
Percent Similarity: 30.12% Conservatives: 82
Best Local Similarity: 19.56% Mismatches: 254
Query Match: 3.71% Indels: 289
DB: 15 Gaps: 38
```

```
US-10-691-383-2 (1-676) x US-10-132-134-1 (1-52101)
Qy 5 AlaAlaAspThrThrArgGly---SerProMetPro-AspThrGlyValLeuArgLeuLe 23
Db 14734 GCAGCAGATGACGACGCGCGCGACCTCGACCGGACGACTTCTACGCGCGCTCGG 14793
Qy 23 uThrSerGluGlnArgAlaLysGlyTyrArgArgGlnLeuGluGlyGluLysSerLeuG1 43
Db 14794 CACCGCGGGCTCGACCGCGCGGTACGCGCGAGGTGCGAAGGGTCCACCGAATCGA 14853
Qy 43 yPheHis-----ProSer-----GluThrProTyrIleLe 53
Db 14854 CGCCACCGGCTGCTCGTACGGGTGCGCGAACCGGCATGTCGACGCGCGACAGCA 14913
Qy 53 sTyrLeuGluGlySerGluThrTyrLysLysValLysLeuProThrAspGlyIleSer-- 72
Db 14914 GCAGTGCAT---CTCCGCGCTGGCGCTCGTGGGCTG---ACCCAGGCTGTTCAGCA 14967
Qy 73 -----AlaSerLysIleLeuGlyLysIleLe 81
Db 14968 CGCGTGGGCGCGCGCGCGCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 15027
Qy 81 tAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPh 101
Db 15028 GGAGCGCACCGG-----GGCATGCTGCTGGCGG----- 15058
Qy 101 eAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGly----- 118
Db 15059 -----ACGTCCGACCGCGCTCTTC-----CATCGCGCTTCTCTCGA 15093
Qy 119 ----GluGlyArgHisLeuGlnThrCysThrAsnSerAsp-----AspAlaLeuAs 134
Db 15094 CGAGGACGCGCGCGGTGCTGGCGCGGTGCGAGCGCGGAGTTCACCGCGCGCGACCTGGA 15153
Qy 134 pProThrAlaProAsnArgAspAsnValAlaPheAla----- 147
Db 15154 GCGGCACTCCCGGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15213
Qy 148 -----SerArgArgAspAlaAlaArgArgG1 156
Db 15214 GGTGCTGGAGACGCGCGTGTGTACGGGGAGGTGGCAGCAGTCCGAGGCGCGTGGCGCGGA 15273
Qy 156 uArgAspGlyThrGlyThrValCysGlnIleThrAsnGly----- 169
Db 15274 GGCGGACCGCTCGGTGACCGTTCGCGCGGTGCGCGGACGGGCGCGCGCGCTGCTGCGCGTC 15333
```

```
QY 170 -----GluThrAspLeuAlaTh 175
Db 15334 GCTCGGGAGACCGTCCGACCTGCTCAAGTTCCGACCTGGCGGACATCGACCTCGACAC 15393
QY 175 rMetPheHisLys---SerLeuProHisAspGluLeuGlyGlnValThrAlaAspAspPh 194
Db 15394 GCATCTCCACGCGTACGGCTTCAGAGTCCATCGCGCTGGCCAACTGGCCTCGGAACCTCAA 15453
QY 194 eAlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAspValProAl 214
Db 15454 CGCGCTCTCGGACGACCTCACCCCGCGCTTCTTCGAGTGTCTCCGAC-----15505
QY 214 aGlyAspProAlaGlyArgLeuValAsnProThrAlaAlaPheAlaIleAspIleSerGln 234
Db 15506 -----ATCCGAGCTCGCGGACCTGCTGCTCGACCGCTACGG 15543
QY 234 yProAlaPheSerAlaThrThr-----IleProProValProThrLeuSe 249
Db 15544 CCCGAGCTGAGCTCCCGACGCGCGCGCGCGCGCTCGCGCCCGCCCGCC 15603
QY 249 rSerProGluLeuAlaGlnLeuAlaGluLeuTyTrpMetAlaLeuAlaArgAspVa 269
Db 15604 GTCCCCA-----GT 15612
QY 269 lProPheMetGlnTyTrpGlyThrAspGluIleThrThrAlaAlaAlaAsnLeuAlaG1 289
Db 15613 GCCGATCGCGCACCCCGCGCGGACGAC-----GACGCGTGGCCATGCTCGCGCGC 15663
QY 289 yMetGlyGlyPheProAsnLeuAspAlaValSerIleGlySerAspGlyThrValAspPr 309
Db 15664 TGCCGCGCGGTTCCTCCCGCGCGGAC-----GACCTGGACAC 15699
QY 309 oPheSerGlnLeuPheArg-----AlaThrPheValGly-----Va 321
Db 15700 CTTCTGCGACGAGTGGCGCGCGCGGAGGACCTGATCGCGGACTACCCCGCGCGACCGCTT 15759
QY 321 lGluThrGlyProPheValSerGlnLeuValAsnSerPheThrIleAspAlaIleTh 341
Db 15760 CGACGGGGGCCCCCTACCGGAGGTGTCGCGCGCGGCGGACTTCCGAGGTTCGCGCGCG 15819
QY 341 rValGluProLysGlnGluThrPheAlaProAspLeuAsnTyTrpMetValAspPheAspG1 361
Db 15820 GATCGAG---GGCGTGACCGCTTCGACGCGGACTTCTTCCACCTGTCGCGGTGGAGGC 15876
QY 361 u-----TrpLeuAsnIleGlnAs 367
Db 15877 GGAGCTGATGACCCCGACGACCGGCTGGCCCTGGAGACCGTGTGGGCGCGCTGGAGAA 15936
QY 367 nGlyGly-----ProProAlaGlyProGluGluLeu-----As 378
Db 15937 CGCGCGGTACGCCCGCGCGCGCTCCCGAGAACACCGGGGTCTACTTCGCGGTCTCCGG 15996
QY 378 pGluGluLeuArgPheIleArgAsnAlaArgAspLeuAla-----391
Db 15997 CAGCGACTACCACTGCTCAACGCGAGTGGCGGACCCCGCGGCTTCACCGGCCAC 16056
QY 392 -----ArgValSerPheValAspAsnIleAsn-----400
Db 16057 CGGCAACGCCCACTCGATGCTGGCGCAACCGGATCTCTAGTCTCGACGTGACGCGGCC 16116
QY 401 -----ThrG1 402
Db 16117 GAGCGAACCCGTGACACGCGGCTGCTCGCTGCTCGCTGACCGCGCGCTCGA 16176
QY 402 uAlaTyArg-----GlySerLeuIleLeuLeuG1 412
Db 16177 GCACATCGGTTCGGCGCGATGCGAGATGGCCATCGCGCGGTGTCAACCTGTGCTGAG 16236
QY 412 uLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSerAspArgGlnAl 432
Db 16237 CGTGACACCTTTCGCGCGACGACATGGCGGCG---ATGCTCAGCCCGACGCGCGCTG 16293
```

```
QY 432 aGlyPheValAsnPheGlyThrSerHisTyTrpPheArgLeuIleGly---AlaAlaGluLe 451
Db 16294 CAAGACCTTCTCCCGCGCGCGGACGCTACGCTCCGCTCCGAGGGCGTCCCGCGGTGCT 16353
QY 451 uAlaGlnArgAlaSerCysTyTrpGlnLys-----TrpGlnValHisArgPh 466
Db 16354 GCTCAAGCCGCTCGCCCGCGCGGACGCGGACGCGCATCTGGGGCGCTCGTCCGCGG 16413
QY 466 eAlaArg-----468
Db 16414 CAGCGCGGAGAACACACGCGCGCGCGCGCTTCGCTGACCGCCCCAACGCGAAGCGCA 16473
QY 469 -----ProGluAlaLeuGlyGlyThr-- 475
Db 16474 GCGCGCCCTCATCAGGACGCCATCGCGGCATCGAGCCCGGACGACATCGCTACGTCGA 16533
QY 476 -LeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAs 495
Db 16534 GCGCACGCGCACGCGGACCGCGC---CTGGCGGACCGCGTCAAGCTCAACGCCCTCGACAG 16590
QY 495 nAspGluLeuLeuLysArgValaGluIleAsnAlaAlaGlnAsnProAsnAsnGluVa 515
Db 16591 CGCTACCGCGCGCTCGCACCGCGCGG-----16618
QY 515 lThrTyTrpLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyProSe 535
Db 16619 -----GGCGGCGCGCGCACGCGCGCGCGCTG 16647
QY 535 r-----GlyHisAlaThrGlnAsnGlyAlaPheAlaTh 546
Db 16648 CGCGCTCGGCTCGGTGAAGACCAACATCGGCCACGCGGAGTCCGCGCGCGCTGGCGCG 16707
QY 546 rValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGlyCysPhePro-----562
Db 16708 AGTGTGAAGTGTCTGCTCGCCARG---CGTACCGGAGCTCGCGCGCGCTTGCACGTG 16764
QY 563 -----AsnProValPheProSerAspAspGlyLeuLeuLeuIle 575
Db 16765 CGACCGGCTCAACCCGACCTCGCTCGACGCGGATTCGAGGTCTGTA 16813

RESULT 5
US-10-437-963-29553
; Sequence 29553, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 29553
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_34043C.1
US-10-437-963-29553

Alignment Scores:
Pred. No.: 2,91e-05 Length: 1236
Score: 129.00 Matches: 110
Percent Similarity: 33.83% Conservative: 50
Best Local Similarity: 23.26% Mismatches: 142
Query Match: 3.66% Indels: 172
```



```
QY 63 Lys-----ValLysLeuProThrAspGlyIleSerAlaSerLysIleLeuGlyLys 79
Db 877 GAGATTTCGGCTCTTCGAACCCACAGCCACAGCGTCCCGCGTCTACGAGAAG 936
QY 80 IleMetAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeu 99
Db 937 CTCTACGACACCATCCGCGAGCGCGCC-----966
QY 100 AlaPheAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGlyGlu 119
Db 967 -----GACTCCCGCGTCTCCGAGCGGTCTTC---GAGTGGCGACCCGCGTGGCGCGC 1017
QY 120 GlyArgHisLeuGlnThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsn 139
Db 1018 GCCGCCAC-----ACCACGACACCCCGGATCGGCTGGCGCGCGCAC 1065
QY 140 Arg----ArgAspAsnValAlaPheAlaSerArgAspAlaAlaArgGluArgAsp 158
Db 1066 GCCCTCGCGACACGCTCGTCTCGACATCCGAGCGCCCTTGGGGGGAGATCGAG 1125
QY 159 -----GlyThrGlyThrValCysGlnIleThrAsnGlyGluThrAspLeuAla 174
Db 1126 TTTTTCATCTCCGCGCGCTCGCTG-----TCCCGGAGCTGTGT 1167
QY 175 ThrMetPheHis---LysSerLeuProHisAspGluLeuGlyGlnValThrAlaAspAsp 193
Db 1168 CGCTGTTCACGCATCGGCTGCC-----1194
QY 194 PheAlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAspValPro 213
Db 1195 -----ATCCTCGAAGGC---TACGCGCTCACCGAGACGAGCCCG 1230
QY 214 AlaGlyAspProAlaGlyArgLeuValAsnProThrAlaAlaPheAlaIleAspIleSer 233
Db 1231 GTG-----CTCGGCTCAACCCCTACGAGGACCCCATCGCGCACCATC 1275
QY 234 GlyProAlaPheSerAlaThrIleProValProThrLeuSerSerProGlu---252
Db 1276 GGCCCGCGGTACGGACACCGAGCTACCGTCGCGCGCGCCCAAGTGTTCGAC 1335
QY 253 -----LeuAlaAlaGlnLeuAlaGlu 259
Db 1336 CGCCAGCGCTGTGACGCGCGCGGAGCTGCTCGCGCGCGCCCAAGTGTTCGAC 1395
QY 260 LeuTyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGluIle 279
Db 1396 GGCTACTGG-----GGGCTGCCCGATGCC 1419
QY 280 ThrThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaVal 299
Db 1420 ACCGACCGCGGTTCGTACCCGCGAGGGCAAGAGTGTTCGACACCGCGCGGTGTC 1479
QY 300 SerIleGlySerAspGlyThrVal-----AspProPheSerGlnLeuPheArg---315
Db 1480 GAATCCGCGCGCGAGCGGTACGTGCGGTCTCTCGAAGCGCGCAAGCAACTGCTGACCCCTC 1539
QY 316 AlaThrPheValGlyValGluThrGlyProPheValSerGlnLeuLeuValAsnSerPhe 335
Db 1540 TCGACGGGCAAGACGTGCGCCCGCGCCCATCGAGGACCGGTTCGCGCGCGCGCGCTG 1599
QY 336 ThrIleAspAlaIleThrValGluProLysGlnGluThrPheAlaProAspLeuAsnTyr 355
Db 1600 GTCCACAGCCATGCTGTCGGGAGCGGCGAGAGTGTCTCCGCGATCCTG-----1653
QY 356 MetValAspPheAsp-----GluTrpLeuAsnIleGlnAsnGlyGlyProProAla 372
Db 1654 GTTCCGAACCTTCGACGCGGTCTCGAGTGGCGCCCGACGAGAGATCGCGCTCCCC---1710
QY 373 GlyProGluGluLeuAspGluGluLeuArgPheIleArgAsnAlaArgAspLeuAlaArg 392
Db 1711 -----GACGACCGCGAGCGCATCTGCCGCGACGAGCGCGTCCGCGCGCGC 1755
QY 393 Val----SerPheValAspAsnIleAsnThr-----GluAlaTyrArg-GlySerLeu11 409
```

```
Db 1756 ATCCAGTCGCGGTTCGACGAGCTCACACCGCGTTCGAGCGCTACGAGCGGATCA-----1810
QY 409 eLeuLeuGluLeuGlyAlaPheSerArgPro 419
Db 1811 -----AGCAGTTCGCGCTTCGTCGCGACCG 1834

RESULT 7
US-10-329-079-48
; Sequence 48, Application US/10329079
; Publication No. US20030198981A1
; GENERAL INFORMATION:
; APPLICANT: FARNET, Chris
; APPLICANT: ZAZOPOULOS, Emmanuel
; APPLICANT: STAFPA, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF LIPOPEPTIDES
; FILE REFERENCE: 3002-11US
; CURRENT APPLICATION NUMBER: US/10/329,079
; CURRENT FILING DATE: 2002-12-24
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 48
; LENGTH: 7185
; TYPE: DNA
; ORGANISM: Streptomyces refuineus
US-10-329-079-48

Alignment Scores:
Pred. No.: 0.00186 Length: 7185
Score: 124.50 Matches: 170
Percent Similarity: 29.79% Conservative: 60
Best Local Similarity: 22.02% Mismatches: 240
Query Match: 3.53% Indels: 302
DB: 15 Gaps: 38

US-10-691-383-2 (1-676) x US-10-329-079-48 (1-7185)
QY 1 MetLeuCysHisAlaAlaAspThrThrArgGlySerProMet-----14
Db 3493 CTGCGCGCCACCGCTTCGACCTGGAGCGCAACCCCATCGGGCTGGTCTGGAG 3552
QY 15 -----ProAspThrGlyValLeuArgLeuThrSerGluGlnArgAlaLysGlyTrp 32
Db 3553 ACCGCGCGCACAGCGGGTCTCGTGTGTCGACCATCGCACATCGCAGCGCGGTGG 3612
QY 33 -----ArgArgGlnLeu-----GluGlyGluLys 40
Db 3613 TCGGCGAGGAGGCTCTCTCGCGACCTGTTACCGCTTACACCGCACCGCGCGCGCGG 3672
QY 41 SerLeuGlyPheHisProSerGluThrProTyrIleLysTyrLeuGluGlySerGluThr 60
Db 3673 GCCCGAAGTGGCGCGCTGCGCGGTGTCAGTAGTGAGTAC-----GCCCTG 3720
QY 61 TrpLysLysValLysLeu-----ProThrAspGlyIleSerAlaSerLysIleLeuGly 78
Db 3721 TGGCAGCGCGGTTCCTCGCGCATCCCGGACCCCGCGCAGCAC-----3765
QY 79 LysIleMetAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCys 98
Db 3766 -----GCAGCGCGCCAG 3777
QY 99 LeuAlaPheAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGly 118
Db 3778 CTGAGTACTGGAGCGGACGCTGCGCGGTCTG---CCGAGGAGCTGAGGCTCCCGGCC 3834
QY 119 Glu-GlyArgHisLeuGlnThrCysThrAsnSerAspAlaLeuAspProThr-Alap 138
Db 3835 GACCGCGCGC-----GTCCGCGCTCCCGTCC 3851
QY 138 roAsnArgArgAspAsnValAlaPheAlaSerArgAspAlaAlaArgGluArg 158
Db 3862 CGCACCGCGCGCGAGTCTGCGCTCCCGCATCCCGTCCACACCGCGGTGTCGAC 3921
```

Qy 158 spGlyThrGlyThrValCysGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPheH 178  
 Db 3922 CTGCGCCGCGAGCTGCGCGGCGAGCGTGTTCATGCTCGTCCAGGCGCG-----C 3969  
 Qy 178 isLysSerLeuProHis---Asp-GluLeuGlyGlnValThrAlaAspPheAlaIle 196  
 Db 3970 GTCGACGCTTCTTCACCCCGATGGCGCGCG----- 4002  
 Qy 197 LeuGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAspValProAlaGlyAsp 216  
 Db 4003 -----GAGGACATCCCGTCGGCACC 4023  
 Qy 217 Pro---AlaGlyArg-----LeuValAsnPro 224  
 Db 4024 CCGGTGCGCGCGCGCAGCAGCGCGGTGGAGACCTGTGCGATTCTTCGTCAAC--- 4080  
 Qy 225 ThrAlaAlaPheAlaIleAspIleSerGly---ProAlaPheSerAlaThrThrIlePro 243  
 Db 4081 ACCCTGTCTCTGCGGACGACACTCTCGCGCACCCTCGCGTTCGCC----- 4125  
 Qy 244 ProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMet 263  
 Db 4126 -----GAGCTGTCGACGGGTCCGCGAGACCGCGCTGGCC 4161  
 Qy 264 AlaLeuAla---ArgAspValProPheMetGlnTyrGlyThrAspGluIleThrThr 282  
 Db 4162 GCCTACGCGCAGCAGACCTGCGCTTCGAGCAG-----CTGTTGGAG 4203  
 Qy 283 AlaAlaAlaAsnLeuAlaGlyMetGlyGlyPhePro----- 294  
 Db 4204 CGCCTGAGCCCGCGCGCTGCTCGCGCGCACCCTCTTCAGTGCCTCTCTCTGC 4263  
 Qy 295 AsnLeuAspAlaValSerIleGlySerAspGlyThrValAspProPheSerGlnLeuPhe 314  
 Db 4264 AACAAACCGGAGGAGCAGCTGGCGCGCAGCGCTCCCGCGCCCG----- 4308  
 Qy 315 ArgAlaThrPheValGlyValGluThrGlyProPheValSerGlnLeuLeuValAsnSer 334  
 Db 4308 ----- 4308  
 Qy 335 PheThrIleAspAlaIleThrValGluProLysGln---GluThrPheAlaProAspLeu 353  
 Db 4309 -----GGACTGCGCGTCAAGCCCCCACCAGGTGGAGACCGCGCGGTGCAAGTTC 4356  
 Qy 354 AsnTyrMetValAspPheAspGluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGly 373  
 Db 4357 GACCTGATGTTACCTTCTCTGAG-----GGCCACGGGAGGACGGCGCGCGCGG 4410  
 Qy 374 ProGlu-----GluLeuAspGluLeuArgPheIleArgAsnAlaArgAspLeu 390  
 Db 4411 ATCGAGACCGCGCTGGAGTACAGCGCGCAGCTCTTCGACAGGAGACCGCGCGAGACCTG 4470  
 Qy 391 AlaArgValSerPheValAspAsnIleAsnThrGluAlaTyrArgGlySerLeuIleLeu 410  
 Db 4471 -----CTGAGGCGGTTCGCGCGGATGCTGCGCGCTC 4500  
 Qy 411 LeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSerAspArg 430  
 Db 4501 TGGGCG-----GCGACCGCGC-----GCCCC----- 4524  
 Qy 431 GlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGlu 450  
 Db 4525 -----ATCGGAGCGCGGAG 4539  
 Qy 451 LeuAla-----GlnArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAla 467  
 Db 4540 CTGCTCGCGCGCGCAGCGCGCACACCGTGTGCGCGCGAGTGGACACCGCCCGCGCGCG 4599  
 Qy 468 ArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeuAlaAsp 487  
 Db 4600 -----GGCCTGGTTCGCGACGCTCCCGGAGATGTTCCAGGAGAGGTTCGCGCGGACT 4650  
 Qy 488 PheAspIleSerLeuLeuGluAsnAspGluLeuLeuLysArgValAlaGluIleAsnAla 507

Db 4651 CCAGACCCCGCGCGTGGAGCAGCGCGCGCGCGGCTGACGTACGCGGAACTCAAGCC 4710  
 Qy 508 AlaGlnAsn-----ProAsnAsnGluVal 515  
 Db 4711 CGGCGCAACCGGTCCGCCAGGTGCTGTCGGGACCGGCTCGGCGCCCGCGCGGTG 4770  
 Qy 516 ThrTyrLeuLeuProGlnAlaIle---GlnValGlySer----- 527  
 Db 4771 GCCCTGTGATGCGCGCTCCCTCGGCGAGTCAACCGCTGCTGGCGGTCTCAAGGCC 4830  
 Qy 528 -----ProThrHisProSerTyrProSerGlyHisAlaThr 539  
 Db 4831 GCGCGCGCTTACGTGCGGTGGACCCCGCGCACCCGAGGAGCGCATCGCTTCATGCTG 4890  
 Qy 540 GlnAsnGlyAlaPheAlaThrValLeuAlaLeuIleGlyLeuAspArgGlyGlyGlu 559  
 Db 4891 CGGACAGCGCGCGCGCTGTGCTGCGCGCGAGTCTGCGCGCGGAGCGCGGAG 4950  
 Qy 560 CysPheProAsnProValPheProSerAspAspGlyLeuGluLeuIleAsnPheGluGly 579  
 Db 4951 ATCGCGCGGTTCGCGTCTGTCGCGACGACGCGCGCGCGCGCGGCGGACCGCG 5010  
 Qy 580 -----AlaCysLeuThrTyrGluGlyGluIleAsnLys-----LeuAlaValAsn 594  
 Db 5011 CCCTGCGCGCGCGACCTCACC---GACGCGGACCGGAAACGCCCTTCAGCGCGCAAC 5067  
 Qy 595 ValAlaPhe-----PheAspGly----- 597  
 Db 5068 GCGGTGACGTGCTTACACTCGGCTCCACGGCGCGCGCGCGCGGCGGTGACCGAG 5127  
 Qy 598 -----GlyArgGlnMetLeuGlyIleHisTyrArg----- 607  
 Db 5128 CACCGCGTCTGCTGCTGCGCGTGGCGGAGCGGTACCGCGGTGCGCGCGCGG 5187  
 Qy 608 -----PheAspGly----- 610  
 Db 5188 AGCGGTGTCGAGTCTGCTGCGCTGCTTCGACGCGCGCGGTGCTGAGTCTGATG 5247  
 Qy 611 -----IleGlnGlyLeuLeu 615  
 Db 5248 GCGTTCGCCACCGGAGGACGCTGCTGCTGCGCGCGCGCGCGGAGGCTG 5307  
 Qy 616 LeuGlyGluThrIleThrValArgThrLeu 625  
 Db 5308 CTCGGCGAGACCATCGCGCGCGCGGATC 5337  
 RESULT 8  
 US-10-329-079-34  
 ; Sequence 34, Application US/10329079  
 ; Publication No. US20030198981A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FARNET, Chris  
 ; APPLICANT: ZAZOPOULOS, Emmanuel  
 ; APPLICANT: STAFFA, Alfredo  
 ; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF LIPOPEPTIDES  
 ; FILE REFERENCE: 3002-11US  
 ; CURRENT APPLICATION NUMBER: US/10/329,079  
 ; CURRENT FILING DATE: 2002-12-24  
 ; NUMBER OF SEQ ID NOS: 66  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 34  
 ; LENGTH: 61944  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces refuineus  
 ; US-10-329-079-34  
 Alignment Scores:  
 Pred. No.: 0.0623 Length: 61944  
 Score: 124.50 Matches: 170  
 Percent Similarity: 29.79% Conservative: 60  
 Best Local Similarity: 22.02% Mismatches: 240  
 Query Match: 3.53% Indels: 302

DB:	15	Gaps:	38
US-10-691-383-2 (1-676) x US-10-329-079-34 (1-61944)			
Qy	1	MetLeuCysHisAlaAlaAspThrThrArgGlySerProMet	14
Db	49566	CTGGCGGCCACCGCTTCGACCTGGAAGCCGACCCCATCGGGCTGGCTGGAG	49625
Qy	15	-----ProAspThrGlyValLeuArgLeuLeuThrSerGluGlnArgAlaLysGlyTyr	32
Db	49626	ACCGGCCCGACAGCCCGGGTCTGCTGTGTGCACCATCGCAGCAGCGGTGG	49685
Qy	33	-----ArgArgGlnLeu-----	40
Db	49686	TCGGGAGGAGAGCTCTTCGCGGACCTGTTACCGCCTACACCGCAGCCCGCGGCGG	49745
Qy	41	SerLeuGlyPheHisProSerGluThrProTyrTyrLeuGlySerGluThr	60
Db	49746	GCCCGAACTGGCGGCGCTGCGGTGCAGTGTGGACTAC-----GCCCTG	49793
Qy	61	TrpLysLysValLysLeu-----ProThrAspGlyLeuSerAlaSerLysIleLeuGly	78
Db	49794	TGGCAGCGCGGTTCCTCGCGATCCCGCGACCCCGCAGCACC-----	49838
Qy	79	LysIleMetAlaArgValArgIleAlaThrAlaLeuAlaValLeuAlaAlaProCys	98
Db	49839	-----GCAGCGGCCAG	49850
Qy	99	LeuAlaPheAspGluValThrAlaSerGlyValPheProGluGlnHisLysThrGly	118
Db	49851	CTGGAGTACTGGGAGCGCAGCTGGCGGTCTG-----CCGAGGAGCTGAGGCTGCCGCC	49907
Qy	119	Glu-GlyArgHisLeuGlnThrCysThrAsnSerAspAlaLeuAspProThr-Alap	138
Db	49908	GACCGCGCGC-----GTCGCGCGTCCCGTCC	49934
Qy	138	roAsnArgArgAspAsnValAlaPheAlaSerArgArgAspAlaAlaArgGluArgA	158
Db	49935	CGACCGCGCGCGGTCTGGCTGACGCTGCCCGCATCCCTCCACCGCGGTGTGCAC	49994
Qy	158	spGlyThrGlyThrValCysGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPheH	178
Db	49995	CTGGCCGGAGCTGCCGGCGAGCGTGTTCATGCTGTCAGCGCGC-----C	50042
Qy	178	isLysSerLeuProHis-----Asp-GluLeuGlyGlnValThrAlaAspPheAlaIle	196
Db	50043	GTCCGACCTTCTCACCCGGATGGCGCGCG-----	50075
Qy	197	LeuGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAspValProAlaGlyAsp	216
Db	50076	-----GAGGACATCCCGTCCGACCC	50096
Qy	217	Pro-----AlaGlyArg-----	224
Db	50097	CCGGTGGCGCGGCACCGACGAGCGGTGGAGACCTGCTCGGATTCCTTCGTCACAC	50153
Qy	225	ThrAlaAlaPheAlaIleAspIleSerGly-----ProAlaPheSerAlaThrIlePro	243
Db	50154	ACCTGTCTCTGGCGGACCGACACCTCCGCGGACCCCGCTCGCC-----	50198
Qy	244	ProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMet	263
Db	50199	-----GACCTGCTGGAGCGGTTCGCGAGACCGCGCTGGCC	50234
Qy	264	AlaLeuAla-----ArgAspValProPheMetGlnTyrGlyThrAspGluIleThrThr	282
Db	50235	GCCTACCGCAGCAGCAGCTGCCCTTCAGCAG-----CTGGTGGAG	50276
Qy	283	AlaAlaAlaAsnLeuAlaGlyMetGlyGlyPhePro-----	294
Db	50277	CGCTGTAGCCCGCGCGCTCGCTCGCGCGGACCCGCTCTTCAGGTGCGCCTCTCTGCG	50336
Qy	295	AsnLeuAspAlaValSerIleGlySerAspGlyThrValAspProPheSerGlnLeuPhe	314
Db	50337	AACAACACCGAGGAGCAGCTGGGGCCCGCAGGGCTCCCGCCCGCC-----	50381
Qy	315	ArgAlaThrPheValGlyValGluThrGlyProPheValSerGlnLeuLeuValAsnSer	334
Db	50381	-----	50381
Qy	335	PheThrIleAspAlaIleThrValGluProLysGln-----GluThrPheAlaProAspLeu	353
Db	50382	-----GGACTCGCGCTCAGCCCGCCACCGAGTGGAGACCGCGCGGTCAAGTTC	50429
Qy	354	AsnTyrMetValAspPheAspGluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGly	373
Db	50430	GACCTGATGTTCACTTCTCTGGAG-----GGCCACGGGAGCAGCGCGCGCGG	50483
Qy	374	ProGlu-----GluLeuAspGluLeuArgPheIleArgAsnAlaArgAspLeu	390
Db	50484	ATCGAGACCGCTCGAGTACAGCGCGACCTTCGACAGGAGACCGCGCGAGCCTG	50543
Qy	391	AlaArgValSerPheValAspAsnIleAsnThrGluAlaTyrArgGlySerLeuIleLeu	410
Db	50544	-----CTGAGCGGTTCGCGCGGATGCTGGCGCTC	50573
Qy	411	LeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSerAspArg	430
Db	50574	TGGGCG-----GCGGACCGCGGC-----GGCCCC-----	50597
Qy	431	GlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGlu	450
Db	50598	-----ATCGAGCGCGGAG	50612
Qy	451	LeuAla-----GlnArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAla	467
Db	50613	CTCTCGCGCGCAGCAGCGGCACCGTCTGCGCGAGTGGAAACGCCACCGCGCGCGC	50672
Qy	468	ArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAsp	487
Db	50673	-----GGCTGTGTCGCGCGCTGCCGAGATGTTCCAGGAGCAGGTGCGCGGACT	50723
Qy	488	PheAspIleSerLeuLeuAsnAspGluLeuLysArgValAlaGluIleAsnAla	507
Db	50724	CCGACGCCCCCGCTGGAGCAGCGCGCGCGGTGACGTACGCCGAACTCAACGCC	50783
Qy	508	AlaGlnAsn-----ProAsnAsnGluVal	515
Db	50784	CGGGCCACCGGCTCGCAGGTGCTGTCTCCGCGCAGCGGTCCGCGCGCGCGGCTG	50843
Qy	516	ThrTyrLeuLeuProGlnAlaIle-----GlnValGlySer-----	527
Db	50844	GCCTGTGTATGCCCGCTCCCTCGCGCAGGTACCGCGGTGCTGGCGGTGCTCAAGGCC	50903
Qy	528	-----ProThrHisProSerTyrProSerGlyHisAlaThr	539
Db	50904	GGCGGCGCTTACGTGCGGTGGACCCCGCGCAGCCCGGAGGAGCGCATCGCTTCATGCTG	50963
Qy	540	GlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGlu	559
Db	50964	CGGACAGCGCCCCCGCTGTCTTGGCGCGCGAGTCTGTCGCGCGGAGCGCGGGAG	51023
Qy	560	CysPheProAsnProValPheProSerAspAspGlyLeuLeuIleAsnPheGluGly	579
Db	51024	ATCGCGGGGTCTCCGTCTCTGTGTCGCGCAGCGCGCGCGCGCGGAGCGCGGAG	51083
Qy	580	-----AlaCysLeuThrTyrGluGlyGluIleAsnLys-----LeuAlaValAsn	594
Db	51084	CCGTCCCGCGCGACCTCACC-----GACGGGAGCGGAAACGCCCTCTGACCGCGGCAAC	51140
Qy	595	ValAlaPhe-----	597
Db	51141	GCGGTGATGCTCTTACACCTCCGGCTCCACGGCGCGCGCGCGCGGTGCTGACCGAG	51200
Qy	598	-----GlyArgGlnMetLeuGlyIleHisTyrArg-----	607

Db 51201 CACCGCGTCTGCTGCTGCGCGCGAGCGGAGCGGTACCCGGTGGCGGCGGC 51260  
 Qy 608 -----PheAspGly----- 610  
 Db 51261 AGCGGTGTGTCAGCTCGCGTCCCTTCGACGGCGCGTGGAGCTGCTGATG 51320  
 Qy 611 -----IleGlnGlyLeuLeu 615  
 Db 51321 GCCTTCGCCACCGAGGAGCGTGGTCTGCGCGACCGCGGCTCTTGGCGGGGAGCTG 51380  
 Qy 616 LeuGlyGluThrIleThrValArgThrLeu 625  
 Db 51381 CTCGGCAGACCATCGCGCGCGCGGATC 51410

## RESULT 9

US-10-282-122A-31609  
 ; Sequence 31609, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Liangsu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari  
 ; APPLICANT: Zyskind, Judith  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John  
 ; APPLICANT: Carr, Grant  
 ; APPLICANT: Yamamoto, Robert  
 ; APPLICANT: Forsyth, R.  
 ; APPLICANT: Xu, H.  
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A  
 ; CURRENT APPLICATION NUMBER: US/10/282,122A  
 ; CURRENT FILING DATE: 2003-02-20  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/230,335  
 ; PRIOR FILING DATE: 2000-09-06  
 ; PRIOR APPLICATION NUMBER: 60/230,347  
 ; PRIOR FILING DATE: 2000-09-09  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/267,636  
 ; PRIOR FILING DATE: 2001-02-09  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 78614  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 31609  
 ; LENGTH: 18930  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas putida  
 ; US-10-282-122A-31609

Alignment Scores:  
 Pred. No.: 0.0104 Length: 18930  
 Score: 124.00 Matches: 147  
 Percent Similarity: 30.99% Conservative: 56  
 Best Local Similarity: 22.44% Mismatches: 226  
 Query Match: 3.51% Indels: 229  
 DB: 13 Gaps: 31

US-10-691-383-2 (1-676) x US-10-282-122A-31609 (1-18930)

Qy 111 ProGluGluHisLysHisThrGlyGluGlyArgHisLeuGln ThrCysThrAsn----- 128  
 |||||  
 Db 573 CTTGATCGCGAGCGGCACAGTGGTGGCGGCACTTCAACGTAGCTGTAATGCGCC 632  
 |||||  
 Qy 129 -----SerAspAspAlaLeuAspProThrAlaProAsnArgArgAspAsnValAlaPh 146  
 |||||  
 Db 633 GCAGATCAACAGCGGAAACCTTGTATGTCACCTGACTGACGCGCCGGTAATGTCG-- 690  
 |||||  
 Qy 146 eAlaSerArgAspAlaAlaArgArgGluArgAspGlyThrGlyThrValCysGlnI 166  
 |||||  
 Db 691 -----GCGCTGGCGCTGTCATGTCACCGGACGCACTGCACCACCTG----- 732  
 |||||  
 Qy 166 eThrAsnGlyGluThrAspLeuAla-----ThrMetPheHisLysSe 180  
 |||||  
 Db 733 -----GCGCGACCGACCTGCTATCAATGAGAGGGTAACACCTTACCGCTCGTGC 785  
 |||||  
 Qy 180 rLeuProHisAspGlu-----LeuGlyGlnValTh 190  
 |||||  
 Db 786 CGAGCCCGCTCGACCGGTTTCCTGGAGCGCAGGTGGCGTATTGCTGGGCATCGCGGT 845  
 |||||  
 Qy 190 rAlaAspAspPheAlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIleCysG 210  
 |||||  
 Db 846 GGCCTG-----GCGAGCGCCAGTTCAGCATACCT 878  
 |||||  
 Qy 210 uAspValProAlaGlyAspProAlaGlyArgLeuValAsnProThrAlaAlaPhAlaI 230  
 |||||  
 Db 879 GCAACCGCCACAAAGCGAC-----GGCCAGGCCCTGGAGATCAGCGCGCAGCGCGC 932  
 |||||  
 Qy 230 eAspIleSerGlyProAlaPheSerAlaThrIleProProVal-----ProThrLe 248  
 |||||  
 Db 933 TGGCAATATCTCCGCAATGTCGAGCATCAGCGG--CCGAGCGTGCACACCCAGATAC 989  
 |||||  
 Qy 248 uSerSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyTrpMetAlaLeuAlaArgAs 268  
 |||||  
 Db 990 CACAGCCCTGAACACCGGACCGCTG-----GCCCTGGCGCGATG 1031  
 |||||  
 Qy 268 pValProPheMetGlnTyGlyThrAspGluIleThr-----ThrAlaAlaA 285  
 |||||  
 Db 1032 CGTCACCTTCACCGCGCGGTGAGCTGTGTCTACCGTGCAGGTGCGCGATGTCGCGG 1091  
 |||||  
 Qy 285 aAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSerIleGlySerAspG 305  
 |||||  
 Db 1092 TAACTCTGATTGGCACCCGT-----GTCTGCGTGGCGATGG 1127  
 |||||  
 Qy 305 yThrValAspProPheSerGlnLeuPheArgAlaThrPheValGlyValGluThrGlyPr 325  
 |||||  
 Db 1128 C-----CTGTTACGCTCACCCTGCTGCGCTGCCGACCC----- 1161  
 |||||  
 Qy 325 oPheValSerGlnLeuValAsnSerPheThrIleAspAlaIleThrValGluProLy 345  
 |||||  
 Db 1162 -----AATGTTGAAGCACTCGACGTGCGTCTGTGGATGCGCG 1199  
 |||||  
 Qy 345 sGlnGluThrPheAlaPro-----AspLeuAsnTyMetValAspPheAspGluTr 362  
 |||||  
 Db 1200 CGGCAACAGCTCTGCCCATTCGATTCGCGCGGACATCATACCCGCTGACGCGAGT 1259  
 |||||  
 Qy 362 pLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluGluLeuAr 382  
 |||||  
 Db 1260 AACCAATATTACAGTTGGTCTGACGG--CTTGGCCCTGAGCGGTGCGCGGAGCGCGGTG 1318  
 |||||  
 Qy 382 gPheIleArgAsnAlaArg-----AspLeuAlaArgValSerPheValAs 397  
 |||||  
 Db 1319 CCACTGTGGAAGTGGCGGATGCCAATGGCATGTGTGATCG--GCACGGGTGCTGGGTGCC 1377  
 |||||  
 Qy 397 pAsnIleAsnThrGluAlaTyArgGlySerLeuIleLeuLeuGluLeuGlyAlaPhSe 417  
 |||||  
 Db 1378 -AAC-----GGCACCTTCTCTGATGACCTCGCCCGCCCGCGC 1412  
 |||||  
 Qy 417 rArgProGlyIleAsnGlyProPheIleAspSerAspArgGlnAlaGlyPheValAsnPh 437  
 |||||  
 Db 1413 ACAGCCGGCGAGCAGTTGAGCTGTGGTGCAGACCGAC---CCAGCGGTAATGCTTCACT 1469

```
Qy 437 eGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGlu----- 450
Db 1470 GGCTACCGAGTACGATGATACACGCGCCCGCAGCGCCGAGCAATCTGGCCAT 1529
Qy 451 -----LeuAlaGlnArgAlaSerCysTyrGlnIlyStrpGlnValHi 464
Db 1530 CGATGCTGATGGCACCACCCTCACCGGTACCGCGCGGTGGCAGCGTGGAGGTGCA 1589
Qy 464 sArgPheAlaArgProGluAlaLeuGlyThrLeuHisAsnThrIleAlaGlyAspLe 484
Db 1590 T-----GACGCCAACCGGCAGCTGATTGGCAGTCCCATTCGCCACGC 1631
Qy 484 uAspAlaAspPheAspIleSerLeu-----LeuGluAsnAspGluLeuLeuIlyAsr 501
Db 1632 CGATGGCAGCTTACGATCGAAGTCCCGGCACAGCCCAAGCGCACTGCTGGACGT 1691
Qy 501 gValAla----- 503
Db 1692 GGTGGCCATCGATGACGGCGCGGTTCTCGTGCCTGCACAGATCACCGCGCCGACAT 1751
Qy 504 -----GluIleAsnAl 507
Db 1752 CACGCGCCCTCGCCGACCGACCGAAGTGGCGGTGAGCGCGGACGCGGTATACCGG 1811
Qy 507 aAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGlnValGlyse 527
Db 1812 CCGTGCAGAACCGGGCAGCAGCTGCGCATCGTGGTGGCGAGCAAGCACTGGGCAC 1871
Qy 527 r-----ProThr-----HisProSerTyrProSerG1 536
Db 1872 TGGCGTGTGGCGCGCGCGGTGTTTCAGCTCAACCTCAACCGCCGACCAAGTTGACGG 1931
Qy 536 yHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuIlyAlaLeuIleGlyLeuAspAr 556
Db 1932 CGAAGTCTCGCAA-----GCCACGGCGACC-----GATGC 1961
Qy 556 gGlyGlyGluCysPheProAsnProValPhe-----ProSerAspAspGlyLeuGlu-- 573
Db 1962 CGTGGCAACACATCGCCAGCAGCGGATGATCGGACCGGATATCGACGGGTGCGATAC 2021
Qy 574 -----LeuIleAsnPheGluGlyAlaCysLeuThrty 584
Db 2022 CACGCCACCGGACCGCCGACTGATGTTGGTATCGGCTCGCGCGCAGCACTCAGTGG 2081
Qy 584 rGluGlyGluIleAsnIlyLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyI1 604
Db 2082 TCGAGGCGCAA----- 2091
Qy 604 eHisTyrArgPheAspGlyIleGlnGlyLeuLeuLeuGlyGluThrIleThrValArgth 624
Db 2092 -----GCCGGTCTACCGTGCAGGTACGTGA 2117
Qy 624 rLeuHisGlnGluLeuMetThrPheAlaGluGluAlaThrPheGluPheArgLeuPhePh 644
Db 2118 CGCGGCGGCAACATC-----CTCGCGAC 2141
Qy 644 rGlyGluValIleIlyLeuPheGlnAspGlyThrPheSerIle 658
Db 2142 CGGACCGCTCGCC-----GCCGATGGCACCCTTCACGTGC 2175
```

## RESULT 10

```
US-10-014-717-1
; Sequence 1, Application US/10014717
; Publication No. US20020192778A1
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goslach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
```

```
; CURRENT APPLICATION NUMBER: US/10/014,717
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US/09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-10-014-717-1
Alignment Scores:
Pred. No.: 0.0985 Length: 68750
Score: 123.50 Matches: 130
Percent Similarity: 31.97% Conservative: 65
Best Local Similarity: 21.31% Mismatches: 234
Query Match: 3.50% Indels: 182
DB: 14 Gaps: 26
```

```
US-10-691-383-2 (1-676) x US-10-014-717-1 (1-68750)
Qy 58 SerGluThrTrp-----LysLysValLysLeu 66
Db 51516 AGTGACACCTGGCTGTGGGACGCGCTGGACGCTGGACGCGCCAGCGCGGAGCGTG 51575
Qy 67 ProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIleMetAlaArgValArgIle 86
Db 51576 CCGCTCGAC-----CTGGTGCTCGGACGCTTCATCGCAAGTGGAGGTC 51620
Qy 87 AlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPheAspGluValThrAla 106
Db 51621 ATGGAGCGCTCGCGGAGGCTACATCATCGGCACCTCCGCATATGGACGCTTCTGCG 51680
Qy 107 SerGlyValPheProGluGluHisLysHisThrGlyGluGlyArgHisLeuGln-ThrCy 126
Db 51681 GCT-----GCTGGAGAGCGTCCACAGATAGACGAGTTG 51713
Qy 126 sThrAsnSerAspAspAlaLeuAspProThrAlaProAsnArgArgAspAsnValAlaPh 146
Db 51714 CTGCTCAGGCTTCAATCTCTGCTGTACGAGAGGTCAAGCGATCATGATGGAAACAC 51773
Qy 146 eAlaSerArgArgAspAlaAlaArgArgGluArgAspGlyThrGlyThrValCysGlnI1 166
Db 51774 CTGTGCGGA-----TCGCACTCCTT-GTAGGGGACGAGAGCATTTTGTGAGCTCTCA 51826
Qy 166 eThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSerLeuProHisAspGluLe 186
Db 51827 GCGCTGCGGAGCGCTGATTGGCGCGGTGCTC-----GAGGAGGC 51868
Qy 186 uGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCys----- 200
Db 51869 CGGAGGGGTGTTCCCGCACCTCCAGTCTTATTGATGTTGCAAGTTTCCCGGGGAACG 51928
Qy 201 -----IleLeu-----AsnGlyAs 205
Db 51929 GCTCGCGGACGATTTGACCGGTAAAGCGCTCGCGCTCGAGATCTCTCTCCCTGGTGCTC 51988
Qy 205 pPheSerIleCysGluAspValProAlaGlyAspPro----- 217
Db 51989 GTTCGATATGGCGAGCGAATCTATCGAGATTCGCCCATCGCCGCTTACTCGAACGCGCAT 52048
Qy 218 -----AlaGlyArgLeuValAsnProThrAlaAlaPheAlaI1 230
Db 52049 CGTGGCGGTGCTGTCGAGTCCGGCGCGGGTGGTAGCACCGCTCGGGAATGTTACGAT 52108
Qy 230 eAspIleSerGlyProAlaPheSerAlaThrThrIleProProValProThrLeuSerSe 250
Db 52109 CTTGAGATCGGAGCAGGACGGCGGACCAACCGCGCGCTCTCCCGGTGTTGCTGCC 52168
Qy 250 rProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAla----- 266
Db 52169 TGACCGGACGAGTACCATTTCCAGCATGTTTCTCGCTCTCTCTCTCTCGCGGAGCA 52228
```



```
QY 267 -----ArgAspValProPheMetGlnTyrGlyThrAspGluLeuThrThrAlaAl 284
Db 52229 AAGATTTCGAGATTATCCATTCCTGAAGTAGTGGCATTCGTGATCGCACCAGGAGCCAGC 52288
QY 284 aAlaAsnLeuAlaGlyMetGlyPheProAsnLeuAspAlaValSerIleGlySerAs 304
Db 52289 TGGCCAGGATACACATCAGAGGTTTGACGTCTGCGCGCCCAATGTC---ATCCA 52345
QY 304 pGlyThrValAspProPheSerGlnLeuPheAlaThrPheValGlyValGluThrG1 324
Db 52346 TGGCACC CGGAT-----ATAAGAGCCAGC----- 52370
QY 324 yProPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIleThrValGluPr 344
Db 52371 -----GCGAAGCGTCTCCTG----- 52385
QY 344 oLysGlnGluThrPheAlaPro---AspLeuAsnTyrMetValAsp-----PheAs 360
Db 52386 -----TCGTGTCGCGCCCGAGGCCTTCTGGTCTGTCGAGGGCACAGGGCATCC 52438
QY 360 pGluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluLeuLeuAspGluG1 380
Db 52439 GATCTGTTGATATACACAG---GGATTGATTGAGGGTGGCAGAGTACGAGATGA 52495
QY 380 uLeuArgPhe-----IleArgAsnAlaArgAsp-----LeuAlaArgVa 393
Db 52496 TCTTCGTATCGACCATCGCTCCTGCTCGTCTGCGACCTGGTGTGACGTCTCTGCGCGGGT 52555
QY 393 lSerPheValAspAsnIle-----AsnThrGluAlaTyrArgGlySe 407
Db 52556 AGGCTTTGCGGCGCGCGTGTGTCGACGCGATCTCCGCGGGGATCTCTCGGACA 52615
QY 407 rLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAs 427
Db 52616 GCAGTGATCTCTCG-----CGGCGCGCGGATAGCAGAGCGCGTGTGA 52663
QY 427 pSerAspArgGlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeuIleG1 447
Db 52664 CAGCTCGGTGAGTCG----- 52679
QY 447 yAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGln----- 462
Db 52680 -GCGACCGAATCGCGCGCGCGTGCAGTACGCGCAGGAATGGCGCGATGGCTCCGCTGA 52738
QY 463 ----ValHisArgPheAla----- 467
Db 52739 CGTCTCATCGATGCGTGTGGAGAGATGTACTTCCACGCGCGCGCGCGCGCAGGT 52798
QY 468 -----ArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGlyAs 483
Db 52799 TTGGTCCACGGTCGATTGCGTACCGGTGGAGGCGCGTTCACGAAGCGCGTCTGCTGAGA 52858
QY 483 pLeu-----AspAlaAspPheAspIleSerLeuLeuLeuAsnAspGluLeuLe 499
Db 52859 TCTGCTCCTGTTGCAAGACACCGCGGCGAGGTGTGGCAGAGGTTTCAGGGGCTCGCGCTGCC 52918
QY 499 uLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLe 519
Db 52919 GCAGCTCAGGCTTCTGCTTTCGCGCGCGGAGCCCGCGGAAGAGTGTGTACGCTTT 52978
QY 519 uProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaTh 539
Db 52979 GGAATGCGCAGCGCAAGACCTATACAGAGGCTCCGCGAGCGCGCTTCTTCTCTCCGCG 53038
QY 539 r-----GlnAsnGlyAlaPheAlaThrValLeuLysAlaLe 551
Db 53039 GGGGGCTTGTGCTGTGATGACACCGGCGGACAGGCGCTCGCTCTGTATCGTGTCT 53098
QY 551 uIleGlyLeuAspArgGlyGluCys 560
Db 53099 GGAAGGG-----CGAGGCGAGCGGTGC 53120
```

## RESULT 11

```
US-10-152-886-44
; Sequence 44, Application US/10152886
; Publication No. US20030064491A1
; GENERAL INFORMATION:
; APPLICANT: ECOPIA BIOSCIENCES INC.
; APPLICANT: Farnet, Chris
; APPLICANT: Staffa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF ENEDIYNE RING
; TITLE OF INVENTION: STRUCTURES
; FILE REFERENCE: 3011-3US
; CURRENT APPLICATION NUMBER: US/10/152,886
; CURRENT FILING DATE: 2002-05-21
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 44
; LENGTH: 5820
; TYPE: DNA
; ORGANISM: Amycolatopsis orientalis
US-10-152-886-44
```

## Alignment Scores:

Pred. No.:	0.00204	Length:	5820
Score:	123.00	Matches:	146
Percent Similarity:	31.73%	Conservative:	59
Best Local Similarity:	22.60%	Mismatches:	249
Query Match:	3.49%	Indels:	192
DB:	13	Gaps:	32

US-10-691-383-2 (1-676) x US-10-152-886-44 (1-5820)

```
QY 4 HisAlaAlaAspThrThrArgGlySerProMetProAspThrGlyValLeuArgLeuLeu 23
Db 1816 CAGCGCCAGCATCTACCGCGCGCGAACCTG---TCCACCGCGCGCGAGGTGCC 1872
QY 24 ThrSerGluGlnArgAlaLys-----GlyTrpArgArgGlnLeuGluGlyGluLys 40
Db 1873 ACCGAGTCGCCCGCGCGCATCGTCACCGGTTGCTGCGCGCGGTGCGGTCTCGAAG 1932
QY 41 SerLeuGlyPheHisProSerGluThrProTyrIleLysTyrLeuGluGlySerGluThr 60
Db 1933 AGCTCGGCATCGAGCGCGCCACGGTC-----ACGGCGCCACAGCTCTCGCGAGCTCACC 1986
QY 61 -----TrpLysLysValLysLeuProThrAspGlyIleSerAlaSerLysIleLeu 77
Db 1987 GCCTGCACCTGGCGCGCGCTCACCGAACGCGAAGTGTCTCAACTGGCCAGATCCGC 2046
QY 78 GlyLysIleMetAlaArgVal----- 84
Db 2047 GCGAAGTGTATGCGACCGCGCAGCGAGCGGGCCATGGCGGCGATCGCGGCCACG 2106
QY 85 ---ArgIleAlaThrAlaLeuAla----- 91
Db 2107 CCAGTGTCCCGAGGCGCTTGGCGCGAGGCGGCGAAGAGTCTCATCGCGGGCTACACGCC 2166
QY 92 -----ValValLeuAlaAlaProCysLeuAlaPheAspGluValThrAla----- 106
Db 2167 CCGAGCAGAGTGTCTCTTCCGAGCCGCGGAGCGGATCGACCGCGTGTGTCGCCGTGCC 2226
QY 107 ----- 108
Db 2227 CGCGCGAGGGGTCTACCGCGCGCGCATCAAGTCTCGCACGCGCTTCCACTCGCGCGCG 2286
QY 109 ValPheProGluGluHisLysHisThrGlyGluGlyArgHisLeuGln----- 124
Db 2287 GTCGTCCCGCGCCCGAGGCGCATGACCGGGGATCTCGCCCGCATCGATTCGCGCGGCTC 2346
QY 125 -----ThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArg 140
Db 2347 GACCGCGCGCTGTCTCTCCACGCTGACCGGTGAGTCTCTGCACCGCGCGGAGACCTCGGC 2406
QY 141 -----ArgAspAsnValAlaPheAlaSerArg---ArgAspAlaAlaArgArgGlu 156
```

```
Db 2407 GATCTGCTGCGCAGCAGGTGGTCTCTCGCGTCCGCTTCGCTGAGCGCGCGGAAGTCC 2466
Qy 157 ArgAspGlyThrGlyThrValCysGlnIleThrAsnGlyGlu 170
Db 2467 GCCGAGCGCAGCAGCAGTGGTGATCGAGGTGGCGCGCGCGCGTGTCTACCGCGCTGCTC 2526
Qy 171 -----ThrAspLeuAlaThrMetPheHisSerLeuProHisAsp 184
Db 2527 GGCACCATCGCGCGCGTACCGGTGCTTTCGATCGACACCGACGCTGAGCTCGCG 2586
Qy 185 GluLeuGlyGlnValThrAlaAspAspPheAlaIle 201
Db 2587 CCGGTGCTGAAGTTCGCGGTGCGCGTTCGCGTTCGCGCGCGCAGCTGGAGACGTCCACG 2646
Qy 202 Leu-----AsnGlyAspPheSerIle----- 208
Db 2647 CTGTTTCAGCGCGCGGTGCTGCGCGCGCTCGCGCGCGCGGAGTTCGTTCTCTCGCC 2706
Qy 209 -----CysGluAspValPro-----AlaGlyAspProAlaGlyArgLeuValAsn 223
Db 2707 AGCCGTGCGAGCGCGCGCTCATCGCGCGCTGCTGCTGACCGCGACCGAGTGGCGGAG 2766
Qy 224 ProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThrThrIlePro 243
Db 2767 CCGCGGAGCGCGCGCGGAAAGCGCTCGGAAAGCGGTGGCAGCAGCACCC----- 2817
Qy 244 ProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeuThrTrpMet 263
Db 2818 ---CTGACACCTGCTGCGC-----AAGCTCGCGTCCGAAAGCGGTGGCGCTCGAA 2868
Qy 264 AlaLeuAlaArgAspVal---ProPheMetGlnTrpGlyThrAspGluIleThr----- 280
Db 2869 GCGGTACCGCGCGACGCGATCCCTCGACGATCTGCACCTTTCGTGTCATCAGCTCGCGC 2928
Qy 281 -----ThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsn 295
Db 2929 CAGCTGTTCAACGAGTGACCGCGCGCTCGCGCGCGCGCTGGAGGTATGCGGAAC 2988
Qy 296 LeuAspAlaValSerIleGlySer---AspGlyThrValAspProPheSerGlnLeuPhe 314
Db 2989 TTCGCGACCGTGTGCTCGCGAACTCGCGGAGATGATCAGCAGCTCGCGCGAGCCGCC 3048
Qy 315 Arg-----AlaThrPheValGlyValGluThrGlyProPheValSer 328
Db 3049 AAACCCCGCGACACACGAGCGCGAGTCCGCGCGTC-----GGCCCGTGG----- 3096
Qy 329 GlnLeuLeuValAsnSerPheThrIleAspAlaIleThrValGluProLysGlnGluThr 348
Db 3097 -----GTCCGCGCGTTTCGCGGTGGAGTACGTC---GTCCGCGCGAAGCCG----- 3138
Qy 349 PheAlaProAspLeuAsnTyrMetValAspPheAspGluTrpLeuAsnIleGlnAsnGly 368
Db 3139 ---TCGCGCGGATCTCGCAGCGGTATCTCCACCGCGAGTGGACGCGCTTCGCGCGCGCC 3195
Qy 369 GlyProProAlaGlyProGluGluLeuAspGluLeuArgPheIleArgAsnAlaArg 388
Db 3196 GGTACACCA-----CTGGCGGAGCGCGTGGCG-----GCG 3225
Qy 389 AspLeuAlaArgValSerPheValAspAsnIleAsnThrGluAlaTyrArgGlySerLeu 408
Db 3226 GCATTGCGCGCGCGCGCGTGGCGGACGGT----- 3255
Qy 409 IleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSer 428
Db 3256 GTCTCTCTGCTGAAACGCGCGAGCGGCTTCGCGCGACGTCGCGCGTGTCTCGTGGCGCG 3315
Qy 429 AspArg-----GlnAlaGlyPhe 434
Db 3316 GCGCGCGGTGCTGGCGCGCGCCCAACGCGCGGTTCGCTGCTGTCGAGCAGCGTCTC 3375
Qy 435 ValAsnPheGlyThrSerHisTyrPheArgLeu----- 445
```

```
Db 3376 GCGCGCTCGCGGTGCGGAAGACGCTCCGCGTGGAGACCCGCTCGGCCCGCACCGCATC 3435
Qy 446 IleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnValHisArg 465
Db 3436 GTGCACTCGCGCACCTCGGCCCG----- 3459
Qy 466 PheAlaArgProGluAlaLeuGlyThrLeuHisAsnThrIleAlaGlyAspLeuAsp 485
Db 3460 ---GTGACCCCGAAGCCCTCGACGCGCGGTG---TCCACTGTGGTCACCGAAGTGGCG 3513
Qy 486 AlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLysArgValAlaGluIle 505
Db 3514 GCGACACCGACTTCAGCGAAGTCCGCTACGACACCGCGGAGTCCGCGACGCGTCCGAAG 3573
Qy 506 AsnAlaAlaGlnAsnPro 511
Db 3574 CTGCGCGCGCTGACCCCG 3591

RESULT 12
US-10-194-163-21/c
; Sequence 21, Application US/10194163
; Publication No. US20020172976A1
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce Carter
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS POLYNUCLEOTIDES
; AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: PALO ALTO
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/194,163
; FILING DATE: 04-Nov-2002
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Basu, Shantanu
; REGISTRATION NUMBER: 43,318
; REFERENCE/DOCKET NUMBER: 529282000101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5995
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 21
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1...2843
; SEQUENCE DESCRIPTION: SEQ ID NO: 21
US-10-194-163-21

Alignment Scores:
Pred. No.: 0.00845 Length: 2843
Score: 122.00 Matches: 141
Percent Similarity: 35.52% Conservative: 113
Best Local Similarity: 19.72% Mismatches: 264
```

Query Match: 3.46% Indels: 198  
DB: 13 Gaps: 38

US-10-691-383-2 (1-676) x US-10-194-163-21 (1-2843)

QY 31 GlyTyrArgGlnLeuGluGlyGlyLeuSerLeuGlyPheHisProSerGluThrPro 50  
DB 2190 GGTGGGGGACAAATGGAGATCGACGATATTGTCTCATCTTCGATCGTCGCGTCGCT 2131  
QY 51 TyrIleLys-----TyrLeuGluGlySer 58  
DB 2130 CGCGTGTACCGGACGTTTGACCGGGGGCTGGATCTGCTCCCTCTTTTGAAAGGCCGT 2071  
QY 59 GluThrTrpLysLysValLysLeu-----Pro 67  
DB 2070 CTGAGCTTCTCTGCTCGTCTGCTGTTCAAAAGCGGAACTCGATGATATCGTTGCGGTCCC 2011  
QY 68 ThrAspGlyIleSerAlaSerLysIleLeuGlyLysIleMetAlaArg-----83  
DB 2010 GATACGGCTCTGAACATTCAAGTTCTCTGATGCTCTGAAAGTAAAGATCCGGAAAG 1951  
QY 84 ---ValArgIleAlaThrAlaLeuAlaValValLeuAlaProCysLeuAlaPheAsp 102  
DB 1950 CCAGCGCGTACGACTCAATATCAATACCATCTTGTCTAGGAAGTGC-----1903  
QY 103 GluValThrAlaSerGlyValPheProGluGluHisLysHisThrGlyGluArgHis 122  
DB 1902 CGGCTCTCTGCTCGGATGTGTACAGACCCCAATC-CGACACATTTGGGATTGCTCGA--1847  
QY 123 LeuGlnThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArgAsp 142  
DB 1846 ---CTTGAATGCCAAAGTAGCTCT-----TTTCCAAAGCGGACAGTAGTCGGGACTGAA 1799  
QY 143 AsnValAla-----PheAlaSerArgArgAspAlaAlaArgArgGluArgAsp 158  
DB 1798 AGCTCAGCTGGCCATCTTCTTTTCTCGAAAAAATACCGGTCTCGATCTCAAGA--1742  
QY 159 GlyThrGlyThrValCysGln-IleThr---AsnGlyGluThrAspLeuAlaThrMetPh 177  
DB 1741 -----ACTGCATGCCGATGTCAGTCTCGTGGCGGACATTTGGGAGTGGCTCGTCT 1691  
QY 177 eHisLysSerLeuProHisAspGluLeuGlyGlnValThrAlaAspPheAlaIleLe 197  
DB 1690 TCATATCTTTGCGGAATAGTCACATCGGTATGCTCCGGA-----1648  
QY 197 uGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAspValProAlaGlyAspPr 217  
DB 1647 ----TACTGTGTTCTCGCCCGGAAACTCTGCTTGCCTGCCATAAGGGTGACGACTT 1592  
QY 217 oAlaGlyArgLeuValAsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPh 237  
DB 1591 GGGCGGACGG---ATCCGTCCCTCC-----GACCTGAAAGGTTTGCTCCC 1550  
QY 237 eSerAlaThrThrIleProValProValProThrLeuSerSerProGluLeuAlaGlnLe 257  
DB 1549 GGCACCTCGATACGATGATATATTGAATCTCTGTGTCGCGAGCTGACAGTCGCTTC 1490  
QY 257 uAlaGluLeuTyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAs 277  
DB 1489 CGGCGAA-----GCGGTCTGCTCGCTGAAATCAGT---ATGAACATATGGC---GA 1448  
QY 277 pGluIleThrThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGly-----292  
DB 1447 TGAACCTACCTTCGAGCGCATAGGTTGGCAGAAATAGAGAAAAGGACGATGCA 1388  
QY 292 -----292  
DB 1387 TGTCAAGCTCCGTTGGATGAGCTTAAGCTCGGAGTCCGACCTACATGCTGGCAGAACG 1328  
QY 293 ----PhePro-----AsnLeuAspAlaValSerIle-----301  
DB 1327 TCTTTTCCGATAAGCTCCCCACTCAAAATGCGCCACTTGGGTGCTATGATATATGGC 1268

QY 302 -GlySerAspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGlyVa 321  
DB 1267 AGCCGCTAACCGCGCGCTGGAC-----GATCTCGCTCGAAACTCGTGTGGAGAC 1217  
QY 321 lGluThrGlyProPheValSerGlnLeuValAsnSerPheThrIleAspAlaIle-- 340  
DB 1216 GGATCGCGAGCTTTGACCATACAG-----GGCAGCATCGGCTCTGATTTCTCTGT 1166  
QY 341 -----ThrValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrMetVa 357  
DB 1165 GCGACTTCCGATGATAGTCTCATCGCATCGAAGAGTGAACATCAAGCTCATGGC 1106  
QY 357 lAspPheAspGluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLe 377  
DB 1105 AGACCGCGACCATGG-----CCTGAACATGT 1079  
QY 377 u-----AspGluLeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPh 395  
DB 1078 GCGGTTGCACTGAAGATGAAGCGCCGACTGGCTGCGCGCAATCTCTCTCGGAAAGTGT 1019  
QY 395 eValAspAsnIleAsnThrGluAlaTyrArgGlySerLeu-----IleLeuLeuGl 412  
DB 1018 GGAGGAGCGGTGCGACCTGACTTTCCGAGGCTATACCTATGAGGACCTGACGATCGA 959  
QY 412 uLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSerAspArgGlnAl 432  
DB 958 CTTACAGGCC-----GACAGGGGCGAGTGTCT 932  
QY 432 aGlyPheValAsnPheGlyThrSerHis---TyrPheArgLeuIleGlyAlaAlaGluLe 451  
DB 931 AGGATCTCTGAATGATGAACGCCCAATGCCATATCCGGCTATCTCTCTCGGAGAGGG 872  
QY 451 uAlaGlnArgAlaSerCysTyrGlnLysTrpGln-----ValHisArgPheAl 467  
DB 871 GCTGCCCTTCTCTCGAGTTCTCCCGGTTTCCGATGGAATCTGACAGCGCACAACTT--814  
QY 467 aArgProGluAlaLeu-----472  
DB 813 -CGTCCGATCGACTTTTGGCGCTTCCGACTGCCATCGCGCGATGCTCTTCTGATTTTC 755  
QY 473 -GlyGlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLe 492  
DB 754 TTCGGAGAATTGTCGGGAGTCGTATTGACGATCTGACAGGAACGATACGAATAGATAC 695  
QY 492 uLeuGlu-----AsnAspGluLeu---LeuLysArgValAlaGluIleAsnAl 507  
DB 694 CTTGGATTGCAAACTCCGACAAAGAACTGCATCTGCACAACTGTAACCTGGCTATGGC 635  
QY 507 aAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuPro-----GlnAlaIleGlnVa 525  
DB 634 AAGGGTGGAGCGACATAGCTCCATTACCTCAGTTCCGCTTCTTGTGAGGAGCACTGCA 575  
QY 525 lGlySerProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAl 545  
DB 574 TGGCAGATTCCACATTTGCAAGCTGCTCG-----ATGCTCGAAGGAGTGGTTGG 524  
QY 545 aThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCysPheProAsnProVa 565  
DB 523 CGCACATCTTCTGCTGACTTCGGGCCATAGGTCCGAGCAAGA-----481  
QY 565 lPheProSerAspAspGlyLeuGluLeuIleAsnPheGluGlyAla-----580  
DB 480 -GTGCCGACTGAGATGTCGCTCTGCGTTTGGCGTTCGACGCTCTTTGCCCCAAGGAAGT 422  
QY 581 -CysLeuThrTyrGlu-----GlyGluIleAsnLysLeuAlaValAsnValAlaPh 597  
DB 421 GTGCAGCTTCTCGGTTGCTTGGCCGAAGCGGATTTCTCAAGATCAGGGCAGCCTA 362  
QY 597 eGlyArg-----GlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLe 615  
DB 361 TTCCGAATGCTTCGAGAGCTCGATCTGCACATCTGATCTGATCCGACAGGCTCATG-----310  
QY 615 uLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeu 629



```
Qy 431 lnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluL 451
Db 2409 CGGTATCTGTCACACAGC---CGTACGACAAACGGCGGTGACCCGGCTGTGGACCGACTTCAG 2465
Qy 451 euAlaGlnArgAlaSerCysTyrGlnLysTyrGlnValHisArgPheAlaArgPro 469
Db 2466 CGGCGCCGCGGGCCACATGC---AGATCTCTGGCGGGCCCGCAGTCGCGGCAAGTCCA 2518

RESULT 14
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.: 428 Length: 9025608
Score: 122.00 Matches: 166
Percent Similarity: 31.57% Conservative: 73
Best Local Similarity: 21.93% Mismatches: 280
Query Match: 3.46% Indels: 240
DB: 15 Gaps: 39

US-10-691-383-2 (1-676) x US-10-156-761-1 (1-9025608)
Qy 4 HisAlaAlaAspThrThrArgGlySerProMetProAspThrGlyValLeuArgLeuLeu 23
Db 498627 CACCTGGCCCGACCCACCGGGGGT---GTCAGACGACCTCGTGTGGACGCTGCTT 498683
Qy 24 ThrSerGluGlnArgAlaLysGlyTyrArgArgGlnLeuGluGlyGluLysSerLeuGly 43
Db 498684 CGGCTCTCCGAAGCGGTTCTGGCGGATCACCGAGGCATACATCTGCCGCGGACCTC--- 498740
Qy 44 PheHisProSerGluThrProTyrIleLysTyrLeuGluGlySerGluThrTyrLysLys 63
Db 498741 -----ACCGAGCGCGGTGTCTCGCCCTATCTC-----GTAGAGAGCTTTCGTCA 498785
Qy 64 VallLysLeuProThrAspGlyIleSerAlaSerLysIleLeu----- 77
Db 498786 GATTCCTCTCCACATCTGCTCTCTCGGGTCTGGGTACTCGCGGTGGACGCGGGCGG 498845
Qy 78 -----GlyLysIleMetAlaArgVal-----ArgIleAlaThrAlaLeu--- 90
Db 498846 TCATGCGGTCTGTCGTCGCGAGGTGTACGGGCGCGACGATCGGTGACGGGTCGCG 498905
Qy 91 -----AlaValValLeuAlaAlaProCysLeu 99
Db 498906 CAGCCAGGAATCCACCGCGCGCGGTGGTGGTCCGCGGTGGTCTCTCCAA----- 498953
```

```
Qy 100 AlaPheAspGluValThrAlaSerGlyValPheProGluGluHisLysHis----- 116
Db 498954 -----CATCAGTAGTGTTCGGTTCACCTTC---GACGATCGTGTCCACGTCGTCCGG 499004
Qy 117 -----ThrGlyGluGlyArgHis-----Leu 123
Db 499005 CGGTGCGCGCGCCGTCACCGTCCCGGGCGCGCGCAGCATGGGCTCGTTCGCCCGTAC 499064
Qy 124 GlnThrCysThrAsnSerAspAlaLeuAspProThr-----Ala 137
Db 499065 GAGCAGGGTGGCGCGCGGATCTCGCGCGCGCCAGTCTCGTTCGAACAGCGGCATATAGCG 499124
Qy 138 ProAsnArgArgAsnValAlaPheAlaSerArgArgAspAlaAlaArgGluArg 157
Db 499125 GCCCATCGCGTGAGTGTTCGCGCGCGGACCTCATGAAGTCCAGCGGCTCGAGTGGCG 499184
Qy 158 AspGlyThrGlyThrValCysGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPhe 177
Db 499185 CTGGGCCATCGGTGAGCACCTGTGTGCGAGGGGACCTTCGCGGTGAAGTGTGCAC 499244
Qy 178 His-----LysSerLeuProHisAspGluLeuGlyGlnValThrAlaAspAsp 193
Db 499245 CATCACACCGCGCGCGCTCCCAT----- 499274
Qy 194 PheAlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAspValPro 213
Db 499275 -----GGCTTCAGGTGTGCGCCACGCCC--- 499298
Qy 214 AlaGlyAspProAla-----GlyArgLeuValAsnProThr 225
Db 499299 GTGGCGCATCCAGCCGCCGAGGAGTAGCCGAGCAGACGCGCGCTTCGCGCGGCACA 499358
Qy 226 AlaAlaPheAlaIleAspIle--SerGlyProAlaPheSerAlaThrThrIlePro--- 243
Db 499359 CTCCCGTACGTCGCGGTGCGAGCGGACACACGACCTCGGCTCGGACGCGGACCGGCTC 499418
Qy 244 -----ProValProThrLeuSerSerProGluLeuAlaAlaG 256
Db 499419 GCCCTTCTCTAGCGCGGGTGGACCGCGCCACACG-TCCAGCGGTTCGCGGAAGGGTG 499477
Qy 256 lnLeuAlaGluLeuTyrTyrMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyT 276
Db 499478 CGGCGAACCGCGCGTACTGTGCGCCCGCGTACGACG-----TACGCGC 499525
Qy 276 hrAspGluIleThrThr-----ThrAlaAlaAlaAsnLeuAlaGlyMetGly--- 291
Db 499526 AGAAGCACATACGCGCGAGGGATGTCGCTTCGCGACACCGCACCGCGCGCGCGCGCC 499585
Qy 292 -----GlyPheProAsnLeuAspAlaValSerIleG 302
Db 499586 GGTCCAGTTCTCGGGCGCTCGAACGACGCGCGGAAGCGCGCTCCATCAGCATCC 499645
Qy 302 LysSerAspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGlyValG 322
Db 499646 CCATGGCTCGCGGGCGCGTCTCGCTTCGCTCGCGTCCGCTAGAGTGGACACGACATGT 499705
Qy 322 luthrGlyProPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIleThrV 342
Db 499706 CGGC-GGGCGGTCTCTCCGCGCGCGCGCTCGCGTGTCT-----C 499743
Qy 342 alGluProLysGlnGluThrPheAla---ProAspLeuAsnTyrMetValAspPheAspG 361
Db 499744 TGGAGCGCGCTCAACAGTGTCTTGGCGAGCTTGGCGGCGTGGGATGTGGAAGCCAGT 499803
Qy 361 luthrLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGlu--- 379
Db 499804 CCGAGGCGGCGGCGACAGTGTGCGCGCGACAGTGTGCGCGACGACGATGTGATCGCGGTG 499863
Qy 380 -----GluLeuArgPheIleArgAsnAlaArgAsp---LeuAlaArgValS 394
Db 499864 AGCGAGTCGAAGCGCGAGTCCAGGAAGTGGCGGTGGCGACGACGTCTCTCGCGCGACGCG 499923
```

```
Qy 394 rPheValAspAsn-----IleAsnThrGluAlaTyrArgGlySerLeuIleLeu- 411
Db 499924 TGCCGAGGACCGCGCTCTGGCGCGGACGCTGTTCCAGCAGCGCTCTCTCTGCTCG 499983
Qy 412 -----GluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheI 426
Db 499984 ACCGGGCGAGAGCGCGAGTCGTCGCCAGCGCGCTCGCGGCT---GGTCCCTCG 500040
Qy 426 leAspSer-----AspArgGlnAlaGlyPheValAsnP 437
Db 500041 CGGACGCGCTCTCGGCTCCAGGACCGCTGTACGTGGCGAGTCGCGCAGCAGCGG 500100
Qy 437 heGlyThrSerHis-----T 442
Db 500101 CTGGCGCTCCCGGTGAAGCGGCGGAGCACTTCTCCAGTCGATCGGCGGACGGTG 500160
Qy 442 yrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpG 462
Db 500161 AGGCGCGCTCTGCTCTGCTCGACGCGCTCGCGCAGCGCGGAGAT----- 500204
Qy 462 lNValHisArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaG 482
Db 500205 -----CGCAGCTCTGGGGG-----CATGACGCCAGGCC 500235
Qy 482 lyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLysArgV 502
Db 500236 CGCGCTCTCGCGC-----CTCTGAGCGCG 500262
Qy 502 alAlaGluIleAsnAlaAlaGlnAsnPro-----AsnAsnGluV 515
Db 500263 TCGTCGAC---CATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 500319
Qy 515 alThrTyrLeuLeuProGlnAlaIleGlnVal-----G 526
Db 500320 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 500379
Qy 526 lySerProThr---HisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheA 545
Db 500380 GCCTGCCACCGCTGCCCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 500439
Qy 545 laThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCysPheProAsnProv 565
Db 500440 TCAGCGCGCTCGAGCAACTCGTCGAGGTGGATGCGCGCGGTGAC----- 500483
Qy 565 alPheProSerAspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyrG 585
Db 500484 -----CTTGAGCGCGCGGCTTC----- 500501
Qy 585 luGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArg-----GlnMetLeuG 603
Db 500502 --GGCGAAGCTCTCGAGACCGCGCTCGGC---CAGCATGAGCAGCGCGTGGCGCGCG 500556
Qy 603 lyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGlu 618
Db 500557 GCGTGCAGCAGCGCGGTGAGCGGTGCGCGGACTCGATCGCGCGCGAG 500603
```

## RESULT 15

```
US-10-437-963-65478/c
; Sequence 65478, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
```

```
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 65478
; LENGTH: 5040
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_66522C.1
US-10-437-963-65478

Alignment Scores:
Pred. No.: 0.00286 Length: 5040
Score: 121.00 Matches: 159
Percent Similarity: 30.40% Conservative: 93
Best Local Similarity: 19.18% Mismatches: 277
Query Match: 3.43% Indels: 301
DB: 17 Gaps: 39

US-10-691-383-2 (1-676) x US-10-437-963-65478 (1-5040)

Qy 26 GluGlnArgAlaLysGlyTyrArgArgGlnLeuGluGlyLysSerLeuGlyPheHis 45
Db 3897 CAAGCAGGCTTGAGTACCTGAGGAACAACAGCAGAGGTTACGATGACAGGGG----- 3844
Qy 46 ProSerGluThrProTyrIleLysTyrLeu-----GluGlySerGluThrTrpLysLys 63
Db 3843 -----CCGGTAATTGATATTGTTGTCATGCGCAGCATGCTGATGTGTGG---AGG 3799
Qy 64 ValLysLeuProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIle----- 80
Db 3798 GTTGCTGTTGATACCCAGGCGCTTGAGGCGCAACAAAATTTGTGGAACACTTGCAGATTTT 3739
Qy 81 -----MetAlaArgValArgIle----- 86
Db 3738 GTTCCATTTGCAAACTACAGACTTGAGCGGAAGTTTGGGATTTTACCAATTAGATGCT 3679
Qy 86 ----- 86
Db 3678 TGCTCTTTTGTAGCAAAACATTTACGATGATGAAACCTTGTAGTATAGTAACGATTGC 3619
Qy 87 -----AlaThrAlaLeuAlaValLeuAlaAla-----ProCysLeuAlaPhe 101
Db 3618 TCTCCTCATGCTACCCATGTTGCGGCGCATTTGCAGCTGCTTTTCATCCT----- 3571
Qy 102 AspGluValThrAlaSerGlyValPhePheGluGluHisLysHisThrGlyGluGlyArg 121
Db 3570 GATGAACCTTGTCTCAATGGAGTTGCACCG-----GGTGCA 3535
Qy 122 HisLeuGlnThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArgArg 141
Db 3534 CAATGATTTCTCTCAAGATTGGCGATACCCGTTTAGGATCAATG----- 3490
Qy 142 AspAsnValAlaPheAlaSerArgArgAspAlaAlaArgGluArgAspGlyThrGly 161
Db 3489 -----GAGACAGGACAGGC 3475
Qy 162 ThrVal-----CysGlnIleThrAsn----- 168
Db 3474 CTGGTTAGGCGCTTGTAGCTGCGAGTAGACCAAAATGTCATCTGATTACATGAGCTAT 3415
Qy 169 GlyGluThrAspLeuAlaThrMetPheHisLysSerLeuProHisAspGluLeuGlyGln 188
Db 3414 GGAGAGCTACCTTCTCTGCTGACTATGCGCAGATTTATT-----GATCTTCTAGT 3364
Qy 189 ValThrAlaAspAspPheAlaIle-----LeuGluAspCysIleLeuAsnGlyAspPhe 206
Db 3363 GAGGTGTGACACAGCATCGTATTATTTATTAGTAGTCTGCGAAACAATGGTCCTGCT 3304
Qy 207 SerIleCysGluAspValProAlaGlyAspProAla-----GlyArgLeu 221
Db 3303 TTAACAACCTCGGTGCTCTCTGGTGGTACTAGTTCAAGCATTTATGCGGTTGGCGCTTAT 3244
Qy 222 ValAsnProThrAlaAla----- 227
```

```
Db 3243 GTCTCACCAGCTATGGCTGCTGGAGCTCACTGTGTGTCGTCACCAAGCACCAGCTGAGGGGATG 3184
Qy 228 -----PheAlaIleAspIleSer 233
Db 3183 GAGTACACATGGTCTAGCCGAGGTCCTACTGCTGACGGGGATCTTGGTGTCTCCATTAGT 3124
Qy 234 GlyProAlaPheSerAlaThrIleProValProThr----- 247
Db 3123 GCACCG-----GGTGGAGCAGTGGCTCTCTGTACCAACAGACACTTCAGTCTCGC 3073
Qy 248 -----LeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeu 260
Db 3072 ATGCTCATGAATGAACCTTCCATGTCAFCACCTCTGCTCGCTCGGTGGGTGTCTCTCT 3013
Qy 261 TyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGluIleThr 280
Db 3012 GTTAGTCCCATGAAGCTGAAGGCATTCATTAGCCCTTATATCTGTGAGGAAGGCAATT 2953
Qy 281 ThrThrAlaAlaAlaAsnLeuAlaGlyMet-----GlyGly 292
Db 2952 GAGAATACAGCTGCATCAATAAGTGTGCTGAGGAGAGCTAAACACAGGGCATGGG 2893
Qy 293 PheProAsnLeuAsp----- 297
Db 2892 CTTTGTGAAGTTGATAGGGCTTTTGAATATGCTCAACAGGCTAAGGAATTGCCACTTGT 2833
Qy 298 -----AlaValSerIleGlySerAspGlyThrValAspProPheSerGlnLeuPheArg 315
Db 2832 TCTTATAGATATCATCAATCAATCAAGTGGC-----AAGCCAACTTCTAACTTAGAGGA 2779
Qy 316 AlaThrPheValGlyValGluThrGlyProPheValSerGlnLeuLeuValAsnSerPhe 335
Db 2778 ATATATCTACGTGGGAGTAAACATGCGCGCAACAAGTGAAGTGAAGTTCCTCA- 2725
Qy 336 ThrIleAspAlaIleThrValGluProLysGlnGluThrPheAlaProAspLeuAsnTyr 355
Db 2724 -----CTAGACCCCAAAATTTTCATGAGGATGCAAGTAACATAGGACAA 2683
Qy 356 MetValAspPheAspGluTrpLeuAsnIleGlnAsnGlyProProAlaGlyProGlu 375
Db 2682 TTGTTCCATTGAGGAGTGGCTGCAGCTGATCC-----ACTGATAGCTGTGTATC 2629
Qy 376 GluLeuAspGluGluLeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPhe 395
Db 2628 AAAATTCCTGAGTACATAAATGGTGACAAACAATGGCGGCACTTTTCAACATTGTTGTAAT 2569
Qy 396 ValAspAsnIleAsnThr-----GluAlaTyr----- 404
Db 2568 CTTGTTAATATCAGCAGCGGCTTTCATATTATGAAGTTTATGGTATAGATTGCAAAAGCA 2509
Qy 405 -----ArgGlySerLeuLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsn 422
Db 2508 CCATGGGTGGACCTATTTCCGGGTACCAATCACTGTAATAAGCCATTGTCTCTATCA 2449
Qy 423 Gly-----ProPheIleAspSerArgGlnAlaGlyPheValAsnPheGlyThrSer 440
Db 2448 GGAGAGCCACCTGCATTGACACTTTCAAAT-----CTTTCCTTTAAGTCAGGT 2401
Qy 441 HisTyrPheArgLeu-----IleGlyAlaAla-----GluLeuAlaGln 453
Db 2400 CACATTGAAGAAGGTTTATAAATGTGCCTATTGGAGCTTCATGGGTGGAAGTTACAATG 2341
Qy 454 ArgAlaSerCysTyrGln----- 459
Db 2340 CGCATCAGCCTTTGATCTACTCCTAGGAGGTTCTTTTGGACACTGTTTCAGATCTGTCCA 2281
Qy 460 -----LysTrpGln-----ValHisArg 465
Db 2280 CTGAAGCGACCTATTAAATGGGAGGCTGTGTCTACTTCTCTTCCACCTTCTCTCAAGAT 2221
Qy 466 PheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeuAsp 485
```

```
Db 2220 TTCAGTTCCTCTGTAGAAAGGTGGCTTAACCTTTGGAGTTGTCCATAGCCAGTTCTGTGGTCT 2161
Qy 486 Ala-----AspPheAspIle----- 490
Db 2160 AGCGGAATTGCTAGCCCATGAACCTACCTGTGTGATTGAGATTGTGTTTCATGGAATA 2101
Qy 490 ----- 490
Db 2100 TCCGTTGATCAGAAAAATAATTGGCTTTGATGGTAGTGAGGCACCTGTGCGTGTGTTGCT 2041
Qy 491 --SerLeuLeuGluAsnAspGluLeuLeuLysArgValAlaGluIleAsnAlaGln 509
Db 2040 AGATCTCTATTAGCATCAGAGAGGCTTGTCTCT--GTGGCACTCTGACAAAGGTCAAG 1984
Qy 510 AsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThr 529
Db 1983 ACCCTTATCGACCGAGTGGAGTCTAACTTTGTTCACTT-----CCTCCCAGC 1936
Qy 530 HisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLys 549
Db 1935 CGTGATAGATTACCATCTGGC-----AAGCAAAATCATAGCACTAACTCTGACT 1888
Qy 550 AlaLeuIleGlyLeuAspArgGlyGlyGluCysPhePro-----Asn 563
Db 1887 TACAATTCAAACTAGAGGATGGTGTGAAATAAAGCCCGCTGTGCCCTACTTAACAAC 1828
Qy 564 ProValPheProSerAspGlyLeuGlu-----LeuIleAsnPheGluGly 579
Db 1827 AGAATATAT-----GATAACAAGTTCGAGTCTCAGTATTATAGAAATATCAGATTCAAAC 1774
Qy 580 AlaCysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArg 599
Db 1773 AAGTGTGTCTACTCTTCTGCTGAGCTATCTCTAATTTATGTCAAACTCTCAAAAGGTGAA 1714
Qy 600 GlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThr 619
Db 1713 TACACGCTGCAATTGTACATAAGGCATGACAATGTGCAG-----TTATT-GGAGAAACT 1661
Qy 620 IleThrValArgThrLeuHisGlnGlu 628
Db 1660 GAAGCAGTTGGTTCATTATTATTCGAGAG 1634
```

Search completed: September 19, 2004, 10:43:49  
Job time : 4388 secs

*This Page Blank (uspto)*



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 19, 2004, 04:00:38 ; Search time 145.41 Seconds  
(without alignments)  
2579.931 Million cell updates/sec

Title: US-10-691-383-2

Perfect score: 3528

Sequence: 1 MLCXAADTRGSPMDTGVL.....SIDGCMGSLVYTGVAQCQA 676

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US10691383/runat\_17092004\_102657\_1766/app\_query.fasta\_1.1230  
-DB=Issued\_Patents\_NA -QPM=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=280 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10691383 @CGN 1 1 93 @runat\_17092004\_102657\_1766 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3528	100.0	2931	3	US-09-151-189-1
2	3528	100.0	2931	4	US-09-596-794-1
3	129.5	3.7	11679	4	US-09-328-352-1377
4	123.5	3.5	68750	3	US-09-335-409-1
5	123.5	3.5	68750	4	US-09-568-102-1
6	123.5	3.5	68750	4	US-09-567-969-1
7	123.5	3.5	68750	4	US-09-568-480-1
8	123.5	3.5	68750	4	US-09-568-486-1
9	123.5	3.5	68750	4	US-09-568-472-1
10	123.5	3.5	68750	4	US-09-567-899-1
11	122	3.5	2843	4	US-09-221-017B-21
12	117.5	3.3	4411529	3	US-09-103-840A-1

13	117	3.3	2208	4	US-09-252-991A-15601	Sequence 15601, A
14	117	3.3	3132	4	US-09-252-991A-15639	Sequence 15639, A
c 15	117	3.3	53500	4	US-09-266-965-76	Sequence 76, Appl
c 16	116.5	3.3	1830121	4	US-09-557-884-1	Sequence 1, Appli
c 17	116.5	3.3	1830121	4	US-09-643-990A-1	Sequence 1, Appli
18	114	3.2	3270	4	US-09-252-991A-3763	Sequence 3763, Ap
19	113.5	3.2	10419	4	US-09-408-020-3	Sequence 3, Appli
20	113.5	3.2	42432	4	US-09-408-020-2	Sequence 2, Appli
c 21	112.5	3.2	1791	4	US-09-252-991A-15622	Sequence 15622, A
c 22	111.5	3.2	1527	4	US-09-252-991A-15698	Sequence 15698, A
c 23	110.5	3.1	1848	4	US-09-252-991A-11564	Sequence 11564, A
c 24	110	3.1	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c 25	110	3.1	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c 26	109.5	3.1	71989	4	US-09-443-501A-2	Sequence 2, Appli
27	108.5	3.1	6948	4	US-09-543-681A-1262	Sequence 1262, Ap
28	108	3.1	25165	4	US-09-453-702B-39	Sequence 39, Appl
29	107.5	3.0	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c 30	107	3.0	1629	4	US-09-252-991A-3705	Sequence 3705, Ap
31	106.5	3.0	1308	4	US-09-252-991A-13571	Sequence 13571, A
c 32	106.5	3.0	4131	4	US-09-252-991A-13773	Sequence 13773, A
33	106.5	3.0	8211	4	US-09-252-991A-13656	Sequence 13656, A
34	106	3.0	4647	4	US-09-252-991A-5730	Sequence 5730, Ap
c 35	106	3.0	9515	1	US-08-920-812-13	Sequence 13, Appl
c 36	106	3.0	9515	1	US-08-920-827-13	Sequence 13, Appl
c 37	106	3.0	9515	1	US-08-921-177-13	Sequence 13, Appl
c 38	106	3.0	9515	1	US-08-362-577C-13	Sequence 13, Appl
c 39	106	3.0	9515	2	US-08-920-828-13	Sequence 13, Appl
c 40	105.5	3.0	3492	4	US-09-489-039A-3882	Sequence 3882, Ap
41	105.5	3.0	10480	4	US-09-732-615-13	Sequence 13, Appl
42	104.5	3.0	1158	4	US-09-252-991A-8509	Sequence 8509, Ap
43	104.5	3.0	1912	3	US-08-868-435-11	Sequence 11, Appl
44	104.5	3.0	1912	4	US-08-744-231-11	Sequence 11, Appl
c 45	104.5	3.0	2019	4	US-09-252-991A-8334	Sequence 8334, Ap

ALIGNMENTS

RESULT 1

US-09-151-189-1  
; Sequence 1: Application US/09151189  
; Patent No. 6233457  
; GENERAL INFORMATION:  
; APPLICANT: Vreeland, Valerie  
; APPLICANT: Ng, Kwan L.  
; TITLE OF INVENTION: The Regents of the University of California  
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses  
; FILE REFERENCE: 023070-087100US  
; CURRENT APPLICATION NUMBER: US/09/151,189  
; CURRENT FILING DATE: 1998-09-10  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 2931  
; TYPE: DNA  
; ORGANISM: Fucus distichus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (228)..(2258)  
; OTHER INFORMATION: vanadium bromoperoxidase  
US-09-151-189-1

Alignment Scores:  
Pred. No.: 0  
Score: 3528.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 100.00%  
DB: 3  
Length: 2931  
Matches: 676  
Conservative: 0  
Mismatches: 0  
Indels: 0  
Gaps: 0

US-10-691-383-2 (1-676) x US-09-151-189-1 (1-2931)

Qy 1 MetLeuCysHisAlaAlaAspThrArgGlySerProMetProAspThrGlyValLeu 20  
|||||

Db 228 ATGCTTTGCCATGCGCGGACACGACAAAGAGGCTCTCTATGCTGACACCGGAGTGCTT 287  
Qy 21 ArgLeuLeuThrSerGluGlnArgAlaValysGlyTyrArgArgGlnLeuGluGlyGluLys 40  
Db 288 CGGTGTCTCATATCAGACGACGCGCTTAAGTTGGAGAGCCGAGTTAGAGGGGGAGAA 347  
Qy 41 SerLeuGlyPheHisProSerGluThrProTyrIleLysTyrLeuGluGlySerGluThr 60  
Db 348 TCACTAGTTTTCATCAAGGAGAGAGCGCTTATATCAAGTACTTGGAGAGGCTCTGAGACT 407  
Qy 61 TrpLysLysValLysLeuProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIle 80  
Db 408 TGAAGAAGGTTAAGCTTCCAAACGACGCGATATCGGCTTCCAAAGATCCTGGGTAAAT 467  
Qy 81 MetAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaProCysLeuAla 100  
Db 468 ATGCCACAGGTCGCGATCGCTACCGCTTGGCTGTGTGATCTGGCCGACCGCTGTGGCA 527  
Qy 101 PheAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGlyGluGly 120  
Db 528 TTCGACGAGTCAAGCCAGTGGTGTTCCTTGAGGAAACACAGCACACCGGGGAGGA 587  
Qy 121 ArgHisLeuGlnThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArg 140  
Db 588 AGACACCTCCAGACCTGTACAAACTCCGACGATCGCTGGATCCGACGCGCGCGAATAGA 647  
Qy 141 ArgAspAsnValAlaPheAlaSerArgArgAlaAlaAlaArgArgGluArgAspGlyThr 160  
Db 648 AGGACAAACAGTAGCTTTGGCTCGCGCGGATGCCCGGAGGAGAACGTGACGGGACA 707  
Qy 161 GlyThrValCysGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSer 180  
Db 708 GGGACTCTCTGCCAATCACTACGAGAACTGATTTGGCTACCATGTTCCACAGTCT 767  
Qy 181 LeuProHisAspGluLeuGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCys 200  
Db 768 CTGCCACACGATGAATGGGACAGGTAACCGCAGACGACTTCGCTATCTCGAGGACTGC 827  
Qy 201 IleLeuAsnGlyAspPheSerIleCysGluAspValProAlaGlyAspProAlaGlyArg 220  
Db 828 ATCTTAAACGGAGATTTTCAGCATTTGGAGAGACGTGCTCGGGAGACCGCGGGGTGCG 887  
Qy 221 LeuValAsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThr 240  
Db 888 CTGCTCAATCTACCGCTGGTGTGCGATCGACATATCCGGTCCCGCATCTCGGCTACG 947  
Qy 241 ThrIleProProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeu 260  
Db 948 ACAATACCCCGGTACCTACCTTCTCTCTGAGCTCGCGCTCAGTTGGCGGAGCTA 1007  
Qy 261 TyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGluIleThr 280  
Db 1008 TACTGGATGGCGCTGGCCAGGAGTACCTTTATGAGATGTCGACCGAGAAATATACC 1067  
Qy 281 ThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyPheProAsnLeuAspAlaValSer 300  
Db 1068 ACTACCGCGGAGCAACCTCGCTGGAATGGAGGCTTCCAAATCTGGAGCGCGTCTCG 1127  
Qy 301 IleGlySerAspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGly 320  
Db 1128 ATAGGGTCGATGATGAGTGGACCCGTTCTCCAGCTCTTCCGAGCGACCTTCGTGGT 1187  
Qy 321 ValGluThrGlyProPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIle 340  
Db 1188 GTTGAAACGGGGCCCTTGTCTCTCAGCTGCTGTGAACAGCTTCCACATCGACGCTATT 1247  
Qy 341 ThrValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPheAsp 360  
Db 1248 ACGTTCGACCGAGCAGGAGACATTCGCCCGCGACTTGAATATATGTCGATTTTGAC 1307  
Qy 361 GluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluGlu 380  
Db 1308 GAATGGCTGAACATTCAGAAATGGTGGACCCCGCGCGCGCGGAGGTTAGACGAAGAG 1367

Qy 381 LeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsn 400  
Db 1368 CTGCGTTTATCGTAAACCGCCGACCTGGCCAGGCTCTCTTCGTGGACAATATCAAC 1427  
Qy 401 ThrGluAlaTyrArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGly 420  
Db 1428 ACCGAAGCTTATCGCGGCTCTCTTATCTTACTTGGAGTGGAGCGCTTCAGACGCGCGT 1487  
Qy 421 IleAsnGlyProPheIleAspSerAspArgGlnAlaGlyPheValAsnPheGlyThrSer 440  
Db 1488 ATCAACGGTCCATTCATCGACAGTATCGGACGCGGCTTCGTCAACTTCGCGACGCTCT 1547  
Qy 441 HisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLys 460  
Db 1548 CACTACTTCAGATTGATAGTGTCCGCGAGCTGCGCAGCGCTCGCTCGTGTACCAAAAG 1607  
Qy 461 TrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIle 480  
Db 1608 TGGCAGGTGCATCGATTGTCACGCCCGGAGGCTTCGCGGGTACCTCCCAACACCATC 1667  
Qy 481 AlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLys 500  
Db 1668 GCGGGGATCTAGATCGACACTTCGACATCTCCCTTCTTGAAATGATGAGCTCTTGAA 1727  
Qy 501 ArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuPro 520  
Db 1728 CGTGTGGCGAGATAATCGGCGCAGATCCCAACACAGAGTCACTTACCTTCTTCCA 1787  
Qy 521 GlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGln 540  
Db 1788 CAAGCTATCCAACTGGGATCGCCAAACGACCCCTTCTTACCCGTCGCGCACCGTACCCA 1847  
Qy 541 AsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCys 560  
Db 1848 AATGGAGCATTTGCCACAGTTCTGAAGGCCCTCATTTGGCCTAGATCGGGAGGTGAGTGC 1907  
Qy 561 PheProAsnProValPheProSerAspAspGlyLeuGluLeuIleAsnPheGluGlyAla 580  
Db 1908 TTCCTTAAACCCGCTGTTCACAGGATGACGGCTCGGAACATACTCACTTCGAAGGGGCA 1967  
Qy 581 CysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGln 600  
Db 1968 TGCCTTACATATCAGGAGAGATCAACAGCTCGCGTCAACGTCGCAATTTGGAGGAGCAG 2027  
Qy 601 MetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThrIle 620  
Db 2028 ATGCTGGGCATCCACTATCGTTTCGACGGTATCCAAAGCCCTACTTCTCGGAGAGACATC 2087  
Qy 621 ThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluGluAlaThrPheGluPhe 640  
Db 2088 ACTGTACGAACACTTCACAGGAGCTGATGACGTTCCCGGAGGAGCCACCTTTGAATTC 2147  
Qy 641 ArgLeuPheThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIleAspGly 660  
Db 2148 CGCTTATTCACCGAGAGGTGATCAAACTTTTCAGGACGGGACATTTCTCATCGATGGA 2207  
Qy 661 AspMetCysSerGlyLeuValTyrThrGlyValAlaAspCysGlnAla 676  
Db 2208 GATATGTGTTCCGGTTTGTGTTTACACTGGCGTGGCGGACTGCCAGGCT 2255

## RESULT 2

US-09-596-794-1

; Sequence 1, Application US/09596794

; Patent NO. 6656715

; GENERAL INFORMATION:

; APPLICANT: Vreeland, Valerie

; APPLICANT: The Regents of the University of California

; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases

; TITLE OF INVENTION: and Their Uses

; FILE REFERENCE: 023070-087110US

; CURRENT APPLICATION NUMBER: US/09/596,794

; CURRENT FILING DATE: 2000-06-19

;; PRIOR APPLICATION NUMBER: US 09/151,189  
;; PRIOR FILING DATE: 1998-09-10  
;; NUMBER OF SEQ ID NOS: 20  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 1  
;; LENGTH: 2931  
;; TYPE: DNA  
;; ORGANISM: Fucus distichus  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (228)..(2258)  
;; OTHER INFORMATION: vanadium bromoperoxidase  
US-09-596-794-1

Alignment Scores:  
Pred. No.: 0 Length: 2931  
Score: 3528.00 Matches: 676  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-10-691-383-2 (1-676) x US-09-596-794-1 (1-2931)

Qy 1 MetLeuCyHisAlaAlaAspThrArgGlySerProMetProAspThrGlyValLeu 20  
Db 228 ATGCTTTGCCATGCAGCGGACACGACAAAGAGGCTCTCTATGCTGACACCGGAGTGCTT 287  
Qy 21 ArgLeuLeuThrSerGluGlnArgAlaGlyGlyTrpArgArgGlnLeuGluGly 40  
Db 288 CGGTGTCTCATCAGACGACGCGCTAAAGGTTGGAGACGCCAGTTAGAGGGGGAGAA 347  
Qy 41 SerLeuGlyPheHisProSerGluThrProTyrIleLysTyrLeuGluGlySerGluThr 60  
Db 348 TCACTAGTTTTCATCAACGAGAGACGCTTATATCAAGTACTTGGAGGCTCTGAGACT 407  
Qy 61 TrpLysLysValLysLeuProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIle 80  
Db 408 TGGAAAGAGTTAAGCTTCAACGCGCGCATATCGGCTTCCAGATCCTGGGTAAATTT 467  
Qy 81 MetAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAla 100  
Db 468 ATGGCCAGGGTCCGCATCGCTTACCGCTTGGCTGTGTGTACTTGGCCGACCCCTGTTGGCA 527  
Qy 101 PheAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGlyGluGly 120  
Db 528 TTCGACGAGGTACAGCCAGTGGTGTTCCTTGAGGAAACACAGACACACCGGGGAGGA 587  
Qy 121 ArgHisLeuGlnThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArg 140  
Db 588 AGACACCTCCAGACCTGTACAAACTCCGACGATCGCTGGATCCGACGGCGCGCAATAGA 647  
Qy 141 ArgAspLeuValAlaPheAlaSerArgAspAlaAlaArgGluArgAspGlyThr 160  
Db 648 AGGACAACTAGCTTTTGGCTCGCGCGCGATGCCCGCAGGCGAGAACGTGACGGGACA 707  
Qy 161 GlyThrValCysGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSer 180  
Db 708 GGGACTGTCTGCCAAATCACTACGGAGAACTGATTTGGCTACCATGTTCCACAGTCT 767  
Qy 181 LeuProHisAspGluLeuGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCys 200  
Db 768 CTGCCACACGATGAATGGGACAGGTAAACGACGACGACTTCGCTATCTCTCGAGGACTGC 827  
Qy 201 IleLeuAsnGlyAspPheSerIleCysGluAspValProAlaGlyAspProAlaGlyArg 220  
Db 828 ATCTTAAACCGGAGATTTTCAGCATTTGCGAGACGTGCTCGGGAGACCCCGCGGGTCGC 887  
Qy 221 LeuValAsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThr 240  
Db 888 CTGCTCAATCTCTACCGCTGGTTGCCATGCACATATCCGGTCCCGCATTTCTCGGCTACG 947  
Qy 241 ThrIleProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeu 260

Db 948 ACAATACCCCGGTACCTACCTTTCTCTCTGAGCTCGCGCTCAGTTTGGCGGAGCTA 1007  
Qy 261 TyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGluIleThr 280  
Db 1008 TACTGGATGGCGCTGGCCAGGATGTACCTTTATGAGTATGGACCGACGAATTTACC 1067  
Qy 281 ThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSer 300  
Db 1068 ACTACCGCGGACGCAACCTCGCTGGAATGGAGGCTTCCCAATCTGGACCGCGTGTGC 1127  
Qy 301 IleGlySerAspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGly 320  
Db 1128 ATAGGGTCCGATGGTACGGTGGACCCGCTTCTCCAGCTCTTCCGAGCGACCTTGGTGT 1187  
Qy 321 ValGluThrGlyProPheValSerGlnLeuValAsnSerPheThrIleAspAlaIle 340  
Db 1188 GTTGAACCGGGCCCTTTGTCTCTCAGCTGCTGCTGAAACAGCTTCACCATCGAGCTATT 1247  
Qy 341 ThrValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPheAsp 360  
Db 1248 ACGTCAACCGAAGCAGGAGACATTCGCCCGGACTTGAACATATATGTCGATTTTGAC 1307  
Qy 361 GluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluGlu 380  
Db 1308 GAATGGCTGAACATTCAGAAATGGTGGACCCCGCGCGCCCGAAGAGTTAGACGAAGAG 1367  
Qy 381 LeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsn 400  
Db 1368 CTGGCTTTATCCGTAAACCGCGGACCTGGCCAGGCTCTCTTCTGGGACATATATCAAC 1427  
Qy 401 ThrGluAlaTyrArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGly 420  
Db 1428 ACCGAAGCTTATCGCGGTCTCTTATCTACTTCTGAGCTGGGAGCTTTCAGCAGGCCGCT 1487  
Qy 421 IleAsnGlyProPheIleAspSerAspArgGlnAlaGlyPheValAsnPheGlyThrSer 440  
Db 1488 ATCAACGCTCATTCATCGACAGTATCGGCGCGGCTTCGTCAACTTCGCGACGCTCT 1547  
Qy 441 HisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLys 460  
Db 1548 CACTACTTCAAGTTAGTGTGCGCGCAGCTGGCGAGGCTGCTGTTGTACCAAAAG 1607  
Qy 461 TrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIle 480  
Db 1608 TGGCAGGTGCATCGATTTGACGCGCGGAGGCTCTCGGGGCTACCCCTCCACACCATC 1667  
Qy 481 AlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLys 500  
Db 1668 GCGGGGATCTAGATGACAGCTTCGACATCTCCCTTCTTGAAAATGATGAGCTCTTGAAA 1727  
Qy 501 ArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuPro 520  
Db 1728 CGTGTGGCGGAGATAATGCGGCGCAGAAATCCCAACAGAGGTACCTACTCTTCTTCCA 1787  
Qy 521 GlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGln 540  
Db 1788 CAAGCTATCAAGTGGGATCGCAACGACCCCTTCTTACCGCTCCGCGCACCGTACCCAA 1847  
Qy 541 AsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCys 560  
Db 1848 AATGGAGCATTTGCCACAGTCTTGAAGCCCTCATTTGGCCCTAGATCGGGAGGTGAGTGC 1907  
Qy 561 PheProAsnProValPheProSerAspGlyLeuGluLeuIleAsnPheGluGlyVala 580  
Db 1908 TTCCCTAACCCGCTGTTTCCCAAGCATGACGGCTCGGAACATCACTTCCAAAGGGGCA 1967  
Qy 581 CysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGln 600  
Db 1968 TGCCTTACATATGAGGAGAGATCAACAAGCTCGCGTCAACGCTCGCATTTGGGAGGCGAG 2027  
Qy 601 MetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThrIle 620



Qy 466 PheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeuAsp 485  
Db 2215 -----AGCGTACCAGGTAGTACTTAAC 2238  
Qy 486 AlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuLysArgValAlaGluIle 505  
Db 2239 CGGATGCGAGT-----AAGACGATTGATGCTAAAGTAACGTTTACA 2280  
Qy 506 AsnAlaAlaGlnAsnPro-----AsnAsnGluValThrTyrLeuLeu----- 519  
Db 2281 GATCGCGCAGGTAATAGCAGCAGGTGTTAACGATACACAAATATACATCGATACCAC 2340  
Qy 520 ---ProGlnAla-----Ile 523  
Db 2341 GCACCTGATGCACCAAGTAATTAACCGGTTAACGGGACAGACCGCGATTACAGGTACGGCA 2400  
Qy 524 GlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAla 543  
Db 2401 GAGCGTGTTCACAGTAACGTGCTGCTATTCAGATGCGAGTACAA----- 2448  
Qy 544 PheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCysPheProAsn 563  
Db 2449 -----ACAGTTGTTGACAGACCGGAT---GGCACTTGACAGTACCAAC 2490  
Qy 564 ProValPheProSerAspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThr 583  
Db 2491 CCAGTTTAAATGATGCGGAT----- 2511  
Qy 584 TyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGly 603  
Db 2512 -----AAAGTTACAGCAATGCTACAGATCCAGCA---GGCAACCATCATTAACCA 2559  
Qy 604 IleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThrIleThrValArg 623  
Db 2560 GGTACACACTACTGTTGATGAGTGGTCCAAATACCGATGGTGTATTTACCGTTGAT 2619  
Qy 624 ThrLeuHisGlnGluLeuMetThrPheAlaGluAlaThrPheGluPheArgLeuPhe 643  
Db 2620 TCAGTTACAGCTGACAAATGTTGATTAATGTCATCAGAGCATCAGGCAACGTTACTGTT--- 2676  
Qy 644 ThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIleAspGlyAspMetCys 663  
Db 2677 ACAGGTGATTGAAAAACGTTCCGGCAGATGCGCAAAATACAGTGTGCTGTTGTGATC 2736  
Qy 664 SerGlyLeuValTyrThrGlyValAlaAspCysGlnAla 676  
Db 2737 AATGCCACAGGTATCTGCAACTGTAGTAGCACAGCA 2775

## RESULT 4

US-09-335-409-1  
; Sequence 1, Application US/09335409  
; Patent No. 6121029  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyt. Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/335,409  
; CURRENT FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 68750  
; TYPE: DNA  
; ORGANISM: Sorangium cellulosum  
US-09-335-409-1

Alignment Scores:

Pred. No.: 0.704 Length: 68750

Score: 123.50 Matches: 130  
Percent Similarity: 31.97% Conservative: 65  
Best Local Similarity: 21.31% Mismatches: 234  
Query Match: 3.50% Indels: 182  
Db: 3 Gaps: 26  
US-10-691-383-2 (1-676) x US-09-335-409-1 (1-68750)  
Qy 58 SerGluThrTrp-----LysLysValLysLeu 66  
Db 51516 AGTGACACCTGGTGTGGAGCGCGCGCTGGAGCGCGCGCGCGCGCGCGCGGTG 51575  
Qy 67 ProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIleMetAlaArgValArgIle 86  
Db 51576 CCGGTCGAC-----CTGGTGTGGAGCTTCCATGCGAAGTGGAGGTG 51620  
Qy 87 AlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPheAspGluValThrAla 106  
Db 51621 ATGGAGCGCTCGCGCAGCGGTACATCATCGGCACTCTCGCATATGGAACGTCTTTCG 51680  
Qy 107 SerGlyValPheProGluGluHisLysHisThrGlyGlyArgHisLeuGln-ThrCys 126  
Db 51681 GCT-----GCTGGAGCGGTACACATAGACGAGTTG 51713  
Qy 126 sThrAsnSerAspAlaLeuAspProThrAlaProAsnArgAspAsnValAlaPhe 146  
Db 51714 CTGTCAGGCTCAAAATCTCTGCTACAGGAAGGTCTCAAGCGATGATGGAACAC 51773  
Qy 146 eAlaSerArgAspAlaAlaArgArgGluArgAspGlyThrGlyThrValCysGlnIle 166  
Db 51774 CTTGTGCGCA-----TCGGCATCCTT-GTAGGGGACGAGAGCATTTTGTGAGCTCTCA 51826  
Qy 166 eThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSerLeuProHisAspGluLe 186  
Db 51827 GCGCTGCCGAGCGCTGATTGGCGCGGTGCTC-----GAGGAGGC 51868  
Qy 186 uGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCys----- 200  
Db 51869 CGGAGGGTGTTCGCCACCTCCAGTCTTATTGAGTGTGCAAGTTTTCGCGGGAACG 51928  
Qy 201 -----IleLeu-----AsnGlyAs 205  
Db 51929 GCTCGCGGAGTATTGACCGGTAAAGCGTTCGGCTCGAGATCTCTTCCCTGTGTGCTC 51988  
Qy 205 pPheSerIleCysGluAspValProAlaGlyAspPro----- 217  
Db 51989 GTTCGATATGCGGAGCGAATCTATCGAGATTCGCCCATCGCCGTACTCGAACGGCAT 52048  
Qy 218 -----AlaGlyArgLeuValAsnProThrAlaAlaPheAlaIle 230  
Db 52049 CGTGGCGGTGTCGTTCGAGTCGCGCGCGGTGGTAGCACCGTCGGGAATGTTTCAGCAT 52108  
Qy 230 eAspIleSerGlyProAlaPheSerAlaThrThrIleProProValProThrLeuSerSe 250  
Db 52109 CTTGGAGATCGAGACGAGGACGCGCGCGCACCGCGCGCTCTCCCGGTGTGCTGCC 52168  
Qy 250 rProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAla----- 266  
Db 52169 TGACCGGACGAGTACCATTTCCACGATGTTTCCGCTCTTCTCTCGCGGGAGCA 52228  
Qy 267 -----ArgAspValProPheMetGlnTyrGlyThrAspGluIleThrThrAlaAl 284  
Db 52229 AAGATTTTCGAGATTATTCATTCTGAAGTATGTCATTCGTGATTCGACCGAGGAGCAGC 52288  
Qy 284 aAlaAsnLeuAlaGlyMetGlyPheProAsnLeuAspAlaValSerIleGlySerAs 304  
Db 52289 TGGCCAGGATACGCATCATCAGAGGTTTGATCGTCATCTCGCGGCCAATGTC---ATCCA 52345  
Qy 304 pGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGlyValGluThrGl 324  
Db 52346 TGGACCCCGCAT-----ATAAGGCCAGC----- 52370  
Qy 324 yProPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIleThrValGluPr 344

```

; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1

Alignment Scores:
Pred. No.: 0.704 Length: 68750
Score: 123.50 Matches: 130
Percent Similarity: 31.97% Conservative: 65
Best Local Similarity: 21.31% Mismatches: 234
Query Match: 3.50% Indels: 182
DB: 4 Gaps: 26

US-10-691-383-2 (1-676) x US-09-568-102-1 (1-68750)

Qy 58 SerGluThrTrp-----LysLysValLysLeu 66
|||:|||||
Db 51516 AGTCACACCTGGCTGTGGGACGCGCGCTGGACGGTGGACGGCCGACGGCGGAGCGTG 51575
Qy 67 ProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIleMetAlaArgValArgIle 86
|||:|||||
Db 51576 CCGTGTGAC-----CTGTGCTCGGACGCTTCCATGCGAAGTGGGAGGTC 51620
Qy 87 AlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPheAspGluValThrAla 106
|||:|||||
Db 51621 ATGGAGCGCTCGCGGACGGGTATCATCGGCACCTCTCGCATATGTAACGTTCTCTGC 51680
Qy 107 SerGlyValPheProGluGluHisLysHisThrGlyGluGlyArgHisLeuGln-ThrCy 126
|||:|||||
Db 51681 GCT-----GCTGGAGAGCGCTCACAGATAGACGAGTTG 51713
Qy 126 eThrAsnSerAspAlaLeuAspProThrAlaProAsnArgAspAsnValAlaPh 146
|||:|||||
Db 51714 CTGTCAGGCTTCAATCTCTGCTGTACAGAGGTCTCAAGCATGATGATGGAACAC 51773
Qy 146 eAlaSerArgArgAspAlaAlaAlaArgArgGluArgAspGlyThrGlyThrValCysGlnI 166
|||:|||||
Db 51774 CTGTGTGCGA-----TCGGCATCTT-GTAGGGACGGAGAGCATTTTGTGAGCTCTCA 51826
Qy 166 eThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSerLeuProHisAspGluLe 186
|||:|||||
Db 51827 GCCGCTGCCGAGCGCTGATTTGGCGGCGGTGCTC-----GAGGAGGC 51868
Qy 186 uGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCys----- 200
|||:|||||
Db 51869 CGGGAGGGTGTTCGCGACCTCCAGTCCTATTGTAGTGTGCAAGTTTCCGGGGAACG 51928
Qy 201 -----IleLeu-----AsnGlyAs 205
|||:|||||
Db 51929 GCTCGCGGACGATTGACCGGTAAAGCGCTCGGCTCGAGATCTCTCTTCCTGTGGCTC 51988
Qy 205 pPheSerIleCysGluAspValProAlaGlyAspPro----- 217
|||:|||||
Db 51989 GTTCGATATGCGGAGCGAATCTATCGAGATTCGCCCATCGCCGTTACTCGAACCGCAT 52048
Qy 218 -----AlaGlyArgLeuValAsnProThrAlaAlaPheAlaI 230
|||:|||||
Db 52049 CGTGGCGGTGTGCTGAGTCGCGCGCGGGTGTAGCACCCTCGGGAATGTTTCAGCAT 52108
Qy 230 eAspIleSerGlyProAlaPheSerAlaThrThrIleProProValProThrLeuSerSe 250
|||:|||||
Db 52109 CTTGGAGATCGGAGACGGGACGGCGGACACCGCGCGGTCTCTCCCGGTGTGTCTGCC 52168
Qy 250 rProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAla----- 266
|||:|||||
Db 52169 TGACCGGACGAGTACCATTTCCCGATGTTTCTCCGCTCTTCTCTGTCGCGGAGCA 52228
|||:|||||

```

```

; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goslach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568.102

```

RESULT 5

US-09-568-102-1

```
Qy 267 -----ArgAspValProPheMetGlnTyrGlyThrAspGluLeuThrThrAlaAl 284
Db 52229 AAGATTTTCGAGATTATCCATCTCTGAAGTAGTGGCATCTCGATGTCACCGAGGCGCAG 52288
Qy 284 aAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSerIleGlySerAs 304
Db 52289 TGGCCAGGGATACGCACATCAGAGGTTTGAGCGTCATCGTCGCGGCCCAATGTC---ATCCA 52345
Qy 304 pGlyThrValAspProPheSerGlnLeuPhePheAlaThrPheValGlyValGluThrG1 324
Db 52346 TGGCACC CGCGAT-----ATAAGAGCCACG----- 52370
Qy 324 yProPheValSerGlnLeuValAenSerPheThrIleAspAlaIleThrValGluPr 344
Db 52371 -----GCGAAGCGCTCTCCTG----- 52385
Qy 344 oLysGlnGluThrPheAlaPro---AspLeuAsnTyrMetValAsp-----PheAs 360
Db 52386 -----TCGTTGCTCGCGCCGAGGCGCTTCTGTGCTCGGTGCGAGGGCACAGGGCATCC 52438
Qy 360 pGluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluG1 380
Db 52439 GATCTGTTGCATATACACAGC---GGATTGATTGAGGGGTGGCAGAGTACGAGATGA 52495
Qy 380 uLeuArgPhe-----IleArgAenAlaArgAsp-----LeuAlaArgVa 393
Db 52496 TCTTCGTATCGACCATCGCTCTCTCGCTCGGACCTGTGTGACGTCTCTGCGCGCGGT 52555
Qy 393 lSerPheValAspAsnIle-----AsnThrGluAlaTyrArgGlySe 407
Db 52556 AGCTTTTCGCGAGCGCGCTGAGTCTGCCAGCGCAGGATCTCCGCGGGGATCTCCGACA 52615
Qy 407 xLeuIleLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAs 427
Db 52616 GCAGTGATCTCTCG-----CGGCGCGCGGATAGCAGGACCGCTTGTGA 52663
Qy 427 pSerAspArgGlnAlaGlyPheValAenPheGlyThrSerHisTyrPheArgLeuIleG1 447
Db 52664 CAGCTCCGGTGAGTCG----- 52679
Qy 447 yAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGln----- 462
Db 52680 -GCGACCGAATCGCGCGCGCGGTGAGTACGAGCAGGATGGCGCGATCGCTCCGCTGA 52738
Qy 463 ---ValHisArgPheAla----- 467
Db 52739 CGTGTCTCATCGATGCGTGTGGAGAGATGTACTCCAGCGCGCGCGCGCGCGAGGT 52798
Qy 468 -----ArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGlyAs 483
Db 52799 TTGGGTCCACGGTCGATTGCGTACCGTGGAGGCGCTTCACGAAGGCGCTCGCTGGAGA 52858
Qy 483 pLeu-----AspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLe 499
Db 52859 TCTGCTCTCTGTTCAAGACACCGCGCGAGGTGCTGGCAGAGGTTTCAGGCGGTCCGCTGCC 52918
Qy 499 uLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrIleLeu 519
Db 52919 GCAGCTCAGGCTTCTCTTTCGCGCGCGGAGACCCCGCGGAGAGGTGTGTACGCTTT 52978
Qy 519 uProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaTh 539
Db 52979 GGAATGCGCAGCGCAAGACCTATACAGAGGCTCCGCGAGCGCGTCTTCTTCTCCGCG 53038
Qy 539 r-----GlnAsnGlyAlaPheAlaThrValLeuLysAlaLe 551
Db 53039 GGGGGCTTGGCTGCTGTGATGAGACGCGGAGCGGCTCGCTCGTCTGTCTGTCTGCT 53098
Qy 551 uIleGlyLeuAspArgGlyGlyGluCys 560
Db 53099 GGAAGGG-----CGAGCGAGCGCGTGC 53120
```

RESULT 6

```
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; PRIOR FILING DATE: 2000-05-10
; CURRENT APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-969-1
```

```
Alignment Scores:
Pred. No.: 0.704 Length: 68750
Score: 123.50 Matches: 130
Percent Similarity: 31.97% Conservative: 65
Best Local Similarity: 21.31% Mismatches: 234
Query Match: 3.50% Indels: 182
DB: 4 Gaps: 26
```

US-10-691-383-2 (1-676) x US-09-567-969-1 (1-68750)

```
Qy 58 SerGluThrTrp-----LysLysValLysLeu 66
Db 51516 AGTGACACCTGGTGTGGGAGCGCGCTGGAGCGCGCCAGCGCGAGCGGTG 51575
Qy 67 ProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIleMetAlaArgValArgIle 86
Db 51576 CCGTTCGAC-----CTGGTCTCGGAGCTTCCATGCGAAGTGGGAGGTC 51620
Qy 87 AlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPheAspGluValThrAla 106
Db 51621 ATGAGCGCGCTCGCGAGCGGTACATCATCGGCACCTCTCGCATATGGAACGCTCTCTGC 51680
Qy 107 SerGlyValPheProGluGluHisLysHisThrGlyGluGlyArgHisLeuGln-ThrCy 126
Db 51681 GCT-----GCTGGAGAGCGTCCACGATAGACGAGTTG 51713
Qy 126 sThrAsnSerAspAlaLeuAspProThrAlaProAsnArgAspAsnValAlaPh 146
Db 51714 CTGTCAGGCTTCAATCTCTCTGCTACAGGAGGTCTCAAGCATGATGGATGAACAC 51773
Qy 146 eAlaSerArgArgAspAlaAlaArgArgGluA-gAspGlyThrGlyThrValCysGlnIl 166
Db 51774 CTTGTCCGCA-----TCGGCATCCTT-GTAGGGGACGAGAGCATTTTGTGAGCTCTCA 51826
Qy 166 eThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSerLeuProHisAspGluLe 186
Db 51827 GCCGTCGCCGAGCGCTGATTGGCGCGGTGCTC-----GAGGAGGC 51868
Qy 186 uGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCys----- 200
Db 51869 CGGGAGGTGTTCCCGCACCTCCAGTCCTATTGTAGTGTGCAAGTTTCCCGGGGACG 51928
Qy 201 -----IleLeu-----AsnGlyAs 205
Db 51929 GCTCGCGGACGTATTGACCGGTAAAGCGCTCGCGCTCGAGATCTCTTCTCTGCTGGGCTC 51988
Qy 205 pPheSerIleCysGluAspValProAlaGlyAspPro----- 217
Db 51989 GTTTCGATATGGGAGCGGAATCTATCGAGATTGCCCATCGCCGCTTACTCGAACGCGAT 52048
```





```
Db 51827 GCCCTCCGAGGCTGATTGGCGCGGTGCTC-----GAGGAGGC 51868
Qy 186 uGlnValThrAlaAspPheAlaIleLeuGluAspCys----- 200
Db 51869 CGGAGGGTGTCCCGACCTCCAGTCTTATTGAGTGGTGAAGTTGCCGGGAAACG 51928
Qy 201 -----IleLeu-----AsnGlyAs 205
Db 51929 GCTCGGAGCGTATTGACCGGTAAAGCGCTCGCGCTCGAGATCTCTTCCCTGGTGCCTC 51988
Qy 205 pPheSerIleCysGluAspValProAlaGlyAspPro----- 217
Db 51989 GTTCGATATGCGGAGCGAATCTATCGAGATTCGCCCATCGCCGTACTCGAAGCGCAT 52048
Qy 218 -----AlaGlyArgLeuValAsnProThrAlaAlaPheAlaIle 230
Db 52049 CGTCGGCGGTGCTCGTCCAGTCGGCGCGGTGTAGTACCGTTCGGGAATGTTACGCAT 52108
Qy 230 eAspIleSerGlyProAlaPheSerAlaThrThrIleProProValProThrLeuSerSe 250
Db 52109 CTTGGAGATCGGAGCAGGCGGCGGACCAACCGCGCGCTCTCCCGGTGTTCCTGCC 52168
Qy 250 rProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAla----- 266
Db 52169 TGACCGGAGCGAGTACCATTTCCACGATGTTTCTCCGCTCTTCTGCTCGCGGAGCA 52228
Qy 267 -----ArgAspValProPheMetGlnTyrGlyThrAspGluIleThrThrAlaAl 284
Db 52229 AAGATTTCGAGATTATTCATTCTGTAAGTATGGCAATCTGTGATGTCACCGAGGCGAGC 52288
Qy 284 aAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSerIleGlySerAs 304
Db 52289 TGGCAGCGGATACGACATCAGAGGTTTGACGTCTCGTCCGCGCAATGTC---ATCCA 52345
Qy 304 pGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGlyValGluThrGl 324
Db 52346 TGGCAGCGCGCAT-----ATAAGAGCCACG----- 52370
Qy 324 yProPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIleThrValGluPr 344
Db 52371 -----GCGNAGGCTCTCTG----- 52385
Qy 344 oLysGlnGluThrPheAlaPro---AspLeuAsnTyrMetValAsp-----PheAs 360
Db 52386 -----TCGTTGCTCGCGCCGCGAGGCTTCTGCTGCTGCGGCGCACAGGCGCATCC 52438
Qy 360 pGluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluLeuLeuAspGluGl 380
Db 52439 GATCTGTTGATATACACACG---GGATTGATTGAGGGGTGGCAGAAGTACGAAGATGA 52495
Qy 380 uLeuArgPhe-----IleArgAsnAlaArgAsp-----LeuAlaArgVa 393
Db 52496 TCTTCGTATCGACCATCGCTCTGCTGCTCGGACTGTGTGTCAGCTCTCGCCCGGCT 52555
Qy 393 lSerPheValAspAsnIle-----AsnThrGluAlaTyrArgGlySe 407
Db 52556 AGGCTTTGGAGACCGCGTGTCTGCCAGGCGCGGATCTCCGCGGGGATCTCCGACCA 52615
Qy 407 rLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyPropheIleAs 427
Db 52616 GCAGGTATCTCTCG-----CGCGCGCGGCGCATAGCAGGCGCGCTTGTGA 52663
Qy 427 pSerAspArgGlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeuIleGl 447
Db 52664 CAGTCCCGTGAAGTCG----- 52679
Qy 447 yAlaAlaGlnLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGln----- 462
Db 52680 -GCCACCGATCGCGCGCGCGCTGTCAGTACGCGCAGGAATGGCCGATGCTCGCTGA 52738
Qy 463 ----ValHisArgPheAla----- 467
Db 52739 CGTCTTCATCGGATGCGTGGAGAGATGTACTTCCACCGCGCGCGCGGCGGCGAGGT 52798
```

---

```
Qy 468 -----ArgProGluAlaLeuGlyThrLeuHisAsnThrIleAlaGlyAs 483
Db 52799 TTGGTCCACGGTCGATTGGTACCGGTGGAGCGCTTACGAAGCGCTCGCTGGAGA 52858
Qy 483 pLeu-----AspAlaAspPheAspIleSerLeuLeuLeuAsnAspGluLeuLe 499
Db 52859 TCTGCTCTCTTTCGAAGACACCGGCGAGGTCTGTGGAGAGGTTTCAGGGGTCCCGCTGCC 52918
Qy 499 uLysArgValAlaGluLeuAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLe 519
Db 52919 GCAGCTCGAGCTTCTGCTTCGCGCGCGGACCGCGGAGAGTGTGTGTACGCTTT 52978
Qy 519 uProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaTh 539
Db 52979 GGAATGTCAGCGCAAGACCTATACAGAGGCTCCGCGAGCGCGTCTTCTCTCTCGC 53038
Qy 539 r-----GlnAsnGlyAlaPheAlaThrValLeuLeuAlaLe 551
Db 53039 GGGGCTTGGCTCGTGTGATGACACGAGGCGGAGCGGCTCGCTCGTATCGTGTCT 53098
Qy 551 uIleGlyLeuAspArgGlyGlyGluCys 560
Db 53099 GGAAGGG-----CGAGGCGAGGCGTGC 53120

RESULT 8
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1

Alignment Scores:
Pred. No.: 0.704 Length: 68750
Score: 123.50 Matches: 130
Percent Similarity: 31.97% Conservative: 65
Best Local Similarity: 21.31% Mismatches: 234
Query Match: 3.50% Indels: 182
DB: 4 Gaps: 26

US-10-691-383-2 (1-676) x US-09-568-486-1 (1-68750)
Qy 58 SerGluThrTrp-----LysLysValLysLeu 66
Db 51516 AGTGACACCTGGCTGTGGGACGCGCGCTGGAGCGGCGGCGGAGCGCGAGCGTG 51575
Qy 67 ProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIleMetAlaArgValArgIle 86
Db 51576 CCGCTCGAC-----CTGGTGTCTCGGCGAGCTTCCATGCGAAGTGGGAGGTC 51620
Qy 87 AlathrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPheAspGluValThrAla 106
Db 51621 ATGGAGGCGCTCGCGCGAGGCGGTACATCATCGGCACTCTCCGCATATGGAACGTCTTCTGC 51680
Qy 107 SerGlyValPheProGluGluHisLysHisThrGlyGlyArgHisLeuGln-ThrCy 126
```

Db 51681 GCT-----GCTGGAGAGGCTCACAGATAGACGAGTGTG 51713  
Qy 126 sThrAsnSerAspAlaLeuAaspProThrAlaProAsnArgAspAsnValAlaPh 146  
Db 51714 CTCTCAGGCTCAAAATCTCTGTCTACAGAGAGGTCATCAACGATGATGGAACAC 51773  
Qy 146 eAlaSerArgAspAlaAlaArgAGluAGluAspGlyThrGlyThrValCysGlnIl 166  
Db 51774 CTGTCTCGCA-----TCGGCATCCTT-GRAGGAGCGAGAGCATTTTGTGACTCTCA 51826  
Qy 166 eThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSerLeuProHisAspGluLe 186  
Db 51827 GCCGCTCCGAGAGCTGATTTGGCGGCTGCTC-----GAGGAGGC 51868  
Qy 186 uGlyGlnValThrAlaAspAspPheAlaLeuLeuGluAspCys----- 200  
Db 51869 CGGAGAGGTGTCGCCGACCTCCAGTCCTATTATTGAGTGTGTCAGATTTGCCGGGAACG 51928  
Qy 201 -----IleLeu-----AsnGlyAs 205  
Db 51929 GCTCCGCGAGTATTGACCGGTAAAGCTCGCGCTCGAGATCTCTTCCCTGCTGCTC 51988  
Qy 205 pPheSerIleCysGluAaspValProAlaGlyAspPro----- 217  
Db 51989 GTTCGATATGCGGAGCGAATCTATCGAGATTGCGCCATCGCCGTTACTCGAACGCGAT 52048  
Qy 218 -----AlaGlyArgLeuValAsnProThrAlaAlaPheAlaIl 230  
Db 52049 CGTCCGCGGTGTCGTGAGTCGCGCGCGGTGTAGCACCGTCGCGGAATGTTTCAGCAT 52108  
Qy 230 eAspIleSerGlyProAlaPheSerAlaThrIleProValProValProThrLeuSerSe 250  
Db 52109 CTTGGAGATCGGACAGGAGCGGCGGACCGCGCGCTCTCCCGGTGTTCTGCGC 52168  
Qy 250 rProGluLeuAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAla----- 266  
Db 52169 TGACCGGAGGAGTACCATTTACCGATGTTCTCCGCTCTCTCTGCTCGCGGAGCA 52228  
Qy 267 -----ArgAspValProPheMetGlnTyrGlyThrAspGluLeuThrThrAlaAl 284  
Db 52229 AAGATTTTCGAGATATTCATTTCTGAAGTATGCGATTTCTGATGTCGACGAGGAGCAGC 52288  
Qy 284 aAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSerIleGlySerAs 304  
Db 52289 TGCCCGAGGATACCCACATCAGAGGTTTGAGCTCATCTCGCGCCCAATGTC---ATCCA 52345  
Qy 304 pGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGlyValGluThrG1 324  
Db 52346 TCGACCCCGCAT-----ATAAGAGCCACG----- 52370  
Qy 324 yProPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIleThrValGluPr 344  
Db 52371 -----GCGAAGCGTCTCCTG----- 52385  
Qy 344 oLysGlnGluThrPheAlaPro---AspLeuAsnTyrMetValAsp-----PheAs 360  
Db 52386 -----TCGTTGCTCGCGCCGAGCGCTTCTGCTGCTCGAGGGCACAGGCGATCC 52438  
Qy 360 pGluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluLeuLeuAspGluG1 380  
Db 52439 GATCTGTTTCGATATACACAG---GGATTGATTGAGGGGTGGCAGAAGTACGAGATGA 52495  
Qy 380 uLeuArgPhe-----IleArgAsnAlaArgAsp-----LeuAlaArgVa 393  
Db 52496 TCTTCGTATCGACCATCGCTCTCTGCTGCTCGGACCTGTGTGACGCTCTCGCGCGGT 52555  
Qy 393 lSerPheValAspAsnIle-----AsnThrCluAlaTyrArgGlySe 407  
Db 52556 AGGCTTTTCGAGACCGCTGAGTCTGCCAGGCGAGCATCTCCGCGCGGGATCTCTCGGACA 52615  
Qy 407 rLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAs 427

Db 52616 GCACGTGATCTCTCG-----CGCGCGCGCGGCATAGCAGGAGCGCTTGTGA 52663  
Qy 427 pSerAspArgGlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeuIleG1 447  
Db 52664 CAGCTCGGTGAGTCG----- 52679  
Qy 447 yAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGln----- 462  
Db 52680 -GGACCGAATCCCGCGCGCGCTGCAGTACGCGAGGAATGGCCGATGGCTCCGCTGA 52738  
Qy 463 ----ValHisArgPheAla----- 467  
Db 52739 CGTCGTCCATCGGATGCGTTGAGAGAGATGATCTCCACCGCGCGCGCGCAGGT 52798  
Qy 468 -----ArgProGluAlaLeuLeuGlyGlyThrLeuHisAsnThrIleAlaGlyAs 483  
Db 52799 TTGGGTCCACGCTCGATTGCGTACCAGGTGAGGCGGCTTCCAGAAAGCGCTCGCTGAGA 52858  
Qy 483 pLeu-----AspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLe 499  
Db 52859 TCTGCTCTCTTTCGAAGACACCGCGGAGGTCTGGCAGAGGTTTCAGGGGCTCCGCTGCC 52918  
Qy 499 uLysArgValAlaGluLeuAsnAlaGlnAsnProAsnAsnGluValThrTyrLeuLe 519  
Db 52919 GCAGCTCAGGCTTCTGCTTTCGCGCGCGGACCGCGGAAAGAGTGTGTACGCTTT 52978  
Qy 519 uProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaTh 539  
Db 52979 GGAATGCGACGCGAAAGACCTATACAGAGGTCGCGGACGCGCTTCTTCTCTCCGC 53038  
Qy 539 r-----GlnAsnGlyAlaPheAlaThrValLeuLysAlaLe 551  
Db 53039 GCGGGCTTGGCTCGTCTGATGACACAGGCGGAGAGGCTGCGCTCGTATCGCTGCT 53098  
Qy 551 uIleGlyLeuAspArgGlyGlyGluCys 560  
Db 53099 GGAAGGG-----CGAGGCGAGGCGTGC 53120

RESULT 9  
US-09-568-472-1  
; Sequence 1, Application US/09568472  
; Patent No. 6358719  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkie, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/568, 472  
; PRIOR FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 68750  
; TYPE: DNA  
; ORGANISM: Sorangium cellulosum  
US-09-568-472-1

Alignment Scores:  
Pred. No.: 0.704 Length: 68750  
Score: 123.50 Matches: 130  
Percent Similarity: 31.97% Conservative: 65  
Best Local Similarity: 21.31% Mismatches: 234  
Query Match: 3.50% Indels: 182  
DB: 4 Gaps: 26

US-10-691-383-2 (1-676) x US-09-568-472-1 (1-68750)

```
QY 58 SerGluThrTrp-----LysLysValLeuLeu 66
DB 51516 AGTACACCTGCTGGGACGCGCGCTGGAGCGTGGAGCGCGCGAGCGGTG 51575
QY 67 ProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIleMetAlaArgValArgIle 86
DB 51576 CCGGTGCAC-----CTGGTCTGGGAGCTTCATGCGAAAGTGGGAGGTC 51620
QY 87 AlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPheAspGluValThrAla 106
DB 51621 ATGGAGCGCTCGCGCAGCGGTACATCATCGGCACCTCTCGCATATGGAACGCTTCTGCG 51680
QY 107 SerGlyValPheProGluGluHisLysHisThrGlyGluGlyArgHisLeuGln-ThrCys 126
DB 51681 GCT-----GCTGGAGAGCGTCCACACGATAGACGAGTGTG 51713
QY 126 sThrAsnSerAspAlaLeuAspProThrAlaProAsnArgAspAsnValAlaPh 146
DB 51714 CTGCTAGGCTTCAATCTCTGTCTCTACAGNAGGTCTCATCGCATGATGGATGGAACAC 51773
QY 146 eAlaSerArgArgAspAlaAlaAlaArgArgGluArgAspGlyThrGlyThrValCysGlnI 166
DB 51774 CTGTGTCGCA-----TCGCGCATCTT-GTAGGCGACGAGAGCATTTTGTGAGCTCTCA 51826
QY 166 eThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSerLeuProHisAspGluLe 186
DB 51827 GCGCTCGCGGAGCTGATTTGGCGCGGTGCTC-----GAGGAGGC 51868
QY 186 uGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCys----- 200
DB 51869 CGGAGGGGTGTCGCCACCTCCAGTCTTATTGAGTGTGCAAGTTTCCCGGGGAACG 51928
QY 201 -----IleLeu-----AsnGlyAs 205
DB 51929 GCTCGCGGAGTATTGACCGGTAAAGACCTCGCGCTCGAGTCTCTTCCCTGGTGGCTC 51988
QY 205 pPheSerIleCysGluAspValProAlaGlyAspPro----- 217
DB 51989 GTTCGATATGGCGGAGCAATCTATCGAGATTCGCCCATCGCCGTTACTCGAACGCGAT 52048
QY 218 -----AlaGlyArgLeuValAsnProThrAlaAlaPheAlaI 230
DB 52049 CGTCCGCGGTGTCGTGAGTCGCGCGCGGGGTGATGACCGCTCGGGAATGTTCAGCAT 52108
QY 230 eAspIleSerGlyProAlaPheSerAlaThrIleProProValProThrLeuSerSe 250
DB 52109 CTGGAGATCGGACGAGGCGGCGGACACCGCGCGCTCTCCCGTGTGTGCTGCC 52168
QY 250 rProGluLeuAlaAlaGlnLeuAlaGluLeuTyTrpMetAlaLeuAla----- 266
DB 52169 TGACCGGACGAGTACCATTTCACCGATGTTCTCCGCTCTTCTGCTCGCGCGGAGCA 52228
QY 267 -----ArgAspValProPheMetGlnTyThrGlyThrAspGluIleThrThrAlaAl 284
DB 52229 AAGATTTTCGAGATATTCATCTCTGAAAGTATGGCAATTCGTGATGTCGACGAGGCGCAGC 52288
QY 284 aAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAlaValSerIleGlySerAs 304
DB 52289 TGCCAGGGATACGCACATCAGAGTTTGGAGTCATCGTCGCGGCCAATGTC---ATCCA 52345
QY 304 pGlyThrValAspProPheSerGlnLeuPhePheArgAlaThrPheValGlyValGluThrGl 324
DB 52346 TGGACCCCGCAT-----ATAAGAGCCACG----- 52370
QY 324 yProPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIleThrValGluPr 344
DB 52371 -----GCCAAGCGTCTCCTG----- 52385
QY 344 oLysGlnGluThrPheAlaPro---AspLeuAsnTyTrpMetValAsp-----PheAs 360
DB 52386 -----TCGTGCTCGCGCGCGAGCCCTTCTGGTCTGTCGAGGGGCACAGGGCATCC 52438
QY 360 pGluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluGl 380
```

52439 GATCTGTTGATATCACCACG---GGATTGATTGAGGGGTGGCAGAAGTACGAAGATGA 52495  
QY 380 uLeuArgPhe-----IleArgAsnAlaArgAsp-----LeuAlaArgVa 393  
DB 52496 TCTTCGTATCACCATCCGCTCTCTGCTCGACCTGTTGACCTCTCGCGCGGT 52555  
QY 393 lSerPheValAspAsnIle-----AsnThrGluAlaTyArgGlySe 407  
DB 52556 AGCTTTTCGCGCGCGCTGAGTCTGCCAGCGCAGGATCTCCGCGCGGATCTCTCGGACA 52615  
QY 407 rLeuIleLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyPropheIleAs 427  
DB 52616 GCAGCTGATCTCTCG-----CGCGCGCGCGGATAGCAGAGCGCGCTTGTGA 52663  
QY 427 pSerAspArgGlnAlaGlyPheValAsnPheGlyThrSerHisTyPheArgLeuIleGl 447  
DB 52664 CAGCTCCGCTGAGTCG----- 52679  
QY 447 yAlaAlaGluLeuAlaGlnArgAlaSerCysTyTrpGlnLysTrpGln----- 462  
DB 52680 -GCGACCGAATCCGCGCGCGCTGTCAGTACGCGCAGGAATGGCGCGATGCTCCGCTGA 52738  
QY 463 ---ValHisArgPheAla----- 467  
DB 52739 CGTCGTCATCGATGCGTGTGGAGAGATGTACTTCCACGCGCGCGCGCGCAGGT 52798  
QY 468 -----ArgProGluAlaLeuGluGlyThrLeuHisAsnThrIleAlaGlyAs 483  
DB 52799 TTGGGTCCACGCTCGATTGCTGCTACCGGTGAGGCGGCTTCACGAGGCTCCGCTGCC 52858  
QY 483 pLeu-----AspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeu 499  
DB 52859 TCTGCTCTCTTCCGAAGACACCGCGGAGCTGTCGGCAGAGGTTCCAGGGCTCCGCTGCC 52918  
QY 499 uLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyLeuLe 519  
DB 52919 GCAGCTCGAGGCTTCTGCTTTCGCGCGCGGACCGCGGGAAGAGTGTGTACGCTTT 52978  
QY 519 uProGlnAlaIleGlnValGlySerProThrHisProSerTyProSerGlyHisAlaTh 539  
DB 52979 GGAATGCGACGCGCAAGACCTATACAGAGGCTCCGCGAGCGCGCTTCTTCTCTCGC 53038  
QY 539 r-----GlnAsnGlyAlaPheAlaThrValLeuLysAlaLe 551  
DB 53039 GGGGGCTTGGCTCGTCTGATGACACGCGGCGGACAGGCGCTCGCTCTGATCGCTGCT 53098  
QY 551 uIleGlyLeuAspArgGlyGlyCys 560  
DB 53099 GGAAGGG-----CGAGGCGGAGCGGTGC 53120

RESULT 10  
US-09-567-899-1  
; Sequence 1, Application US/09567899  
; Patent No. 6383787  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/567,899  
; CURRENT FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 68750  
; TYPE: DNA



ADDRESSEE: MORRISON & FOERSTER  
 STREET: 755 PAGE MILL ROAD  
 CITY: Palo Alto  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 94304-1018  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: Windows  
 SOFTWARE: FastSEQ for Windows Version 2.0b  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/221.017B  
 FILING DATE: 23-DEC-1998  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PP1182  
 FILING DATE: 31-DEC-1997  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PP1546  
 FILING DATE: 30-JAN-1998  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PP2911  
 FILING DATE: 09-APR-1998  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/AU98/01023  
 FILING DATE: 10-DEC-1998  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Monroy, Gladys H  
 REGISTRATION NUMBER: 32,430  
 REFERENCE/DOCKET NUMBER: 27340-20021.00  
 TELEPHONE: 650-813-5600  
 TELEFAX: 650-494-0792  
 TELEX: 706141  
 INFORMATION FOR SEQ ID NO: 21:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2843 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: circular  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: UNKNOWN  
 ORIGINAL SOURCE:  
 ORGANISM: PORPHYROMONAS GINGIVALIS  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: 1...2843  
 US-09-221-017B-21

Alignment Scores:  
 Pred. No.: 0.00536 Length: 2843  
 Score: 122.00 Matches: 141  
 Percent Similarity: 35.52% Conservative: 113  
 Best Local Similarity: 19.72% Mismatches: 264  
 Query Match: 3.46% Indels: 198  
 DB: 4 Gaps: 38

US-10-691-383-2 (1-676) x US-09-221-017B-21 (1-2843)

Qy 31 GlyTrpArgGlnLeuGluGlyGluLysSerLeuGlyPheHisProSerGluThrPro 50  
 |||||  
 Db 2190 GGTGGTGGGACAAAGGAGATCGACGATATTGTCATCTTCGATCGCCGCGTCCT 2131  
 |||||  
 Qy 51 TyrIleLys-----TyrLeuGluGlySer 58  
 |||||  
 Db 2130 GCGGTGTGACGCGAAGCTTTGACCGGGGGTGGATCTGCTCCCTCTTTTGAAGGCGGT 2071  
 |||||  
 Qy 59 GluThrTrpLysValLysLeu-----Pro 67  
 |||||  
 Db 2070 CTGAGCTTCTCCTGCTCGTCTGTTTCAAGGGCAACTCGATGATATATCGTTCCGTCCTC 2011  
 |||||

Qy 68 ThrAspGlyIleSerAlaSerLysIleLeuGlyLysIleMetAlaArg----- 83  
 |||||  
 Db 2010 GATACGGCTCTGAACATTCAAGTTCTCTGAGTCTCTGAAAGTAAGAAATCCGGAAGAG 1951  
 |||||  
 Qy 84 ---ValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPheAsp 102  
 |||||  
 Db 1950 CCGAGCCGGCTACGACTCAATATATACATCTTGTCTCAGGAACCTGC----- 1903  
 |||||  
 Qy 103 GluValThrAlaSerGlyValPheProGluGluHisLysHisThrGlyGluGlyArgHis 122  
 |||||  
 Db 1902 CGGCTCTCTGCTCGGAGTTGTACAGACACCAATC-CGACACATTTGGATGCTCGA--- 1847  
 |||||  
 Qy 123 LeuGlnThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArgAsp 142  
 |||||  
 Db 1846 ---CTTGAATGCCAAAGTAGTGTCT-----TTTCCAAGCGGACAGTACGGGACTGAA 1799  
 |||||  
 Qy 143 AsnValAla-----PheAlaSerArgArgAspAlaAlaArgArgGluArgAsp 158  
 |||||  
 Db 1798 AGCTCAGCTGGCCCATCTTCTTTTCTCGAAAAAATACCGGTCTCGATCTCAAGA--- 1742  
 |||||  
 Qy 159 GlyThrGlyThrValCysGln-IleThr---AsnGlyGluThrAspLeuAlaThrMetPh 177  
 |||||  
 Db 1741 -----ACTGCATGCCGATGTCACTCTCTGTCGCGCACACTTTGGGAGTGGCTCGTCT 1691  
 |||||  
 Qy 177 eHisLysSerLeuProHisAspGluLeuGlyGlnValThrAlaAspPheAlaIleLe 197  
 |||||  
 Db 1690 TCATATCTCTTTGCCGAATAGTCACATCGGTATGTCTCCGGA----- 1648  
 |||||  
 Qy 197 uGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAspValProAlaGlyAspPr 217  
 |||||  
 Db 1647 ----TACTGTGTTCTCTGCCCGGAAACTCTGCTTCCTGCCATAGGGTGGAGCGGACTT 1592  
 |||||  
 Qy 217 oAlaGlyArgLeuValAsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPh 237  
 |||||  
 Db 1591 GGGCGGACGG--ATCCGTCCTCTCC-----GACCTGAAAGTTTGTCTCC 1550  
 |||||  
 Qy 237 eSerAlaThrThrIleProProValProThrLeuSerSerProGluLeuAlaAlaGlnLe 257  
 |||||  
 Db 1549 GGCACCTCGATACGATGATATATTGAATCTCTCTGTGCCCGAGCTGACAGTCGCTTC 1490  
 |||||  
 Qy 257 uAlaGluLeuTyrrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrrGlyThrAs 277  
 |||||  
 Db 1489 CGGCGAA-----GCGGTTCTGCTGAAATCAGT---ATGAACATATGCG---GA 1448  
 |||||  
 Qy 277 pGluIleThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGly----- 292  
 |||||  
 Db 1447 TGAAGTACCTTCGACGCGCATAGGTTGGCAGAAATAGGAAAAAGGACCGATGCA 1388  
 |||||  
 Qy 292 ----- 292  
 |||||  
 Db 1387 TGTCAAGACTCCGGTTGGATGAGCTTAAGCTCCGAGTCCGGACCTACATCTGTCGAGAACG 1328  
 |||||  
 Qy 293 ----PhePro-----AsnLeuAspAlaValSerIle----- 301  
 |||||  
 Db 1327 TCTTTTTCGCGATAAGCTCCCACTCAATATGCGCACCTGGGTCTATCGATATATATGCG 1268  
 |||||  
 Qy 302 -GlySerAspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGlyVa 321  
 |||||  
 Db 1267 AGCGCTAACGCGCGCTGCAC-----GATCTCGCTCGAAACTCGTGTGGAGAC 1217  
 |||||  
 Qy 321 lGluThrGlyProPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIle-- 340  
 |||||  
 Db 1216 GGATCGCGAGCTTGTACCATACAG-----GGCAGCATCGCCCTCGATCTCTCTG 1166  
 |||||  
 Qy 341 -----ThrValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrrMetVa 357  
 |||||  
 Db 1165 GCGACTTGGCGATGTAGATGCTCATGCGATACCGAAAGACTGGAACCTCAAGCTCATGCG 1106  
 |||||  
 Qy 357 lAspPheAspGluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLe 377  
 |||||  
 Db 1105 AGACCGGACGATGG-----CTGACATGT 1079  
 |||||  
 Qy 377 u-----AspGluLeuLeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPh 395  
 |||||



Db 2381258 AGCGCGACCGTTCGTCGTCGCCATCAACCGTTCGCGTATTGCGTAACCCACGCTGCC 2381317  
Qy 268 AspValProPheMetGlnTyrGlyThrAspGluIleThrThrAlaAlaAlaAsnLeu 287  
Db 2381318 ACCTGGCCACCGCAGCAGCGCGCGACGAA-----ACGCGCGCGCGGGTATACG 2381371  
Qy 288 AlaGlyMetGlyGlyPheProAsnLeuAsp-----AlaValSerIleGlySerAspGly 305  
Db 2381372 TCCCATTTGGGGGATGCTTACCTACCTACCGAGTTGGCGGCAACCATGTCATCGACGCG 2381431  
Qy 306 ThrValAspProPheSerGlnLeuPheArgAlaThrPheValGlyValGluThrGlyPro 325  
Db 2381432 GCT-----CTGGTACACCACTTCTTCGGTGTCAACACCATCCCG 2381473  
Qy 326 Phe---ValSerGlnLeuValAsnSerPheThrIleAspAlaIleThrValGluPro 344  
Db 2381474 ATCGCCCTCAACGAGCGCGACTACCTGCGCATGTGGATCCAGCGCGCGACCGTGC----- 2381527  
Qy 345 LysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPheAspGluTyrLeuAsn 364  
Db 2381528 -----ATGAGCCACTATCAAGCGCTGCGCGCAC 2381554  
Qy 365 IleGlnAsnGlyGlyProPro-----AlaGlyProGluGluLeuAspGluGluLeu 381  
Db 2381555 GA-AAGCGTGGCGCGACCGCCACGACGCGCGCGCGCGCATAGTAC---CAGTGC 2381610  
Qy 382 ArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsnThr 401  
Db 2381611 GGCCAGCTCGCGGCTAGCAGCAGCTTCCCGACCGCGCAAAATTGATCTCGCAGCTACT 2381670  
Qy 402 GluAlaTyrArgGlySerLeuIleLeu-----LeuGluLeuGlyAlaPheSer 417  
Db 2381671 CAAGGATTTCTCGAGCTGTGCGCTATCTGGCTGTTGAGCTGTGCGCGCGCGCTCGG 2381730  
Qy 418 ArgProGlyIleAsnGlyProPheIleAspSerAspArgGlnAlaGlyPheValAsnPhe 437  
Db 2381731 CGACCTCATCGCCAGGT-----GTTGAGCTGGTTTCATCTC 2381766  
Qy 438 -----GlyThrSerHisTyrPheArgLeuIleGlyAla----- 448  
Db 2381767 GTTCGTGTCGGTCCAGTCTTCAGTTCCTCGCTACTCTGTGCTGCGCGGCTGACCGGCT 2381826  
Qy 449 -----AlaGluLeuAla 452  
Db 2381827 TTTCGACCGTTCGCCCGCTGACAGTCCGCTCTGCTGCTGCGCGGCTGACCGGCT 2381886  
Qy 453 GlnArgAla-----SerCysTyrGlnLys 460  
Db 2381887 TGCCGGGCTCGGTGCGGTATCGGGCGCGCGGACCAATGTCGTAACGTGTGCACTCCGA 2381946  
Qy 461 TrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIle 480  
Db 2381947 TGGTCCC--ACCGCGCAAGCTGCGCTCGCGGCAACCGAGTCACTCTGGTGGGTACC--- 2382001  
Qy 481 AlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLys 500  
Db 2382001 ----- 2382001  
Qy 501 ArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuPro 520  
Db 2382002 -----AACCGGCTCGCTCGGTATACACCGCGCGCACCC 2382034  
Qy 521 GlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGln 540  
Db 2382035 GCTCCGACCACTCGCGGCGCACCGACCGGACCGTCTGCTCCGATCCCGATCCAGTCCGCGCCAA 2382094  
Qy 541 AsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCys 560  
Db 2382095 GGC-----CTTTACGCGTGGTGTCTCCGAC-----GGGGAAGGG 2382130  
Qy 561 PheProAsnProVal 565  
Db 2382131 TTC---AACCCGATC 2382142

## RESULT 13

US-09-252-991A-15601  
; Sequence 15601, Application US/09252991A  
; Patent No. 6551795

## GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 15601

; LENGTH: 2208

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-15601

Alignment Scores:  
Pred. No.: 0.0127 Length: 2208  
Score: 117.00 Matches: 144  
Percent Similarity: 32.99% Conservative: 82  
Best Local Similarity: 21.02% Mismatches: 255  
Query Match: 3.32% Indels: 206  
Gaps: 33  
DB:

US-10-691-383-2 (1-676) x US-09-252-991A-15601 (1-2208)

Qy 52 IleLysTyrLeuGluGly-----SerGluThrTrpLysLys 63  
Db 92 GTCGATTACTTGACGCGCGTGGACCGGATCGTACCGGCGCACCTACCGGCG 151  
Qy 64 ValLysLeuProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIleMetAlaArg 83  
Db 152 ACCAACCGACCGCCTACGCGCTCGCGCACCAAGATCTCAGCACCATCAAGAGCAAG 211  
Qy 84 -----ValArgIleAlaThrAlaLeuAlaValValLeu 94  
Db 212 GAGTACAAAGCGAGTGGCGGCAACGAGCTCGCATCGACACCGCGCGCATCAGT 271  
Qy 95 AlaAlaProCysLeuAlaPheAspGluValThrAlaSerGlyValPhe----- 110  
Db 272 GCGGCG-----CTGATGAGCGACCGGTCGCGGCGGCTGACCTGGGCTACCTA 322  
Qy 111 -----ProGluGluHisLysHisThrGlyGluGlyArgHisLeuGlnThrCys 126  
Db 323 ACCCATCTCGGCGCGGAGCGGCAAGCGCGCGGCGGAGGCTTCGAACCTGCGCACCC-- 379  
Qy 127 ThrAsnSerAspAlaLeuAspProThrAlaProAsnArgArgAspAsnValAlaPhe 146  
Db 380 -----GACGACATGCGCGCTGCGTGGCGGCAAGGG--GCTGCT 417  
Qy 147 AlaSerArgArgAspAlaAla-----ArgArgGluArgAsp--GlyThrGlyThrValCy 164  
Db 418 GCTGAGCACCAGGAGCAGCTCAGGCGCGCGCGCGGCATCTCGACCGAGCGCTGTGT 477  
Qy 164 sGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPhe-----His 178  
Db 478 GCAGGTGCTGAGGCGCGCTGGAGCTGGCGCGGAGCTGCGGAGCTACGCGCGGAGCA 537  
Qy 178 sLysSerLeuProHisAspGluLeuGlyGlnValThrAlaAspAspPheAlaIleLeuGln 198  
Db 538 CCAGGCGCTGGGCGCACGACGCGCGCGCGCAAAACC-----CTCCA 579  
Qy 198 uAspCysIleLeuAsnGlyAspPheSerIleCysGluAspVal-----ProAlaGlyAs 216  
Db 580 GGAAGCGCTG-----CGCGACCTCGGCGCACCGTGCACGCA 615





Db 1558 GCGGCG-----CTGATGAGCGACACCGTGCCAGCGCGTGCACCTGGGGCTACTTA 1608  
Qy 111 -----ProGluGluHiLysHisThrGlyGluGlyArgHisLeuGlnThrCys 126  
Db 1609 ACCATCTCGGCCCGGAGCGGCAAGCGCGCGCGGCGGCTTCGAACTGGCGACC-- 1665  
Qy 127 ThrAsnSerAspAlaLeuAspProThrAlaProAsnArgArgAspAsnValAlaPhe 146  
Db 1666 -----GACGAACATGGCGGTGGCTGGCGGCAAGGG--GCTGCT 1703  
Qy 147 AlasrArgArgAspAlaAa-----ArgArgGluArgasp--GlyThrGlyThrValCy 164  
Db 1704 GCTGAGCACCAGAGAGAGCTCAGGCGCGCGCGCGGCATCTCGACCGAGCGTGTGGT 1763  
Qy 164 sGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPhe-----His 178  
Db 1764 CGAGGTCTGGAGCGCGCGTGGAGCTGGCCGCGAGCTACGCGCGGGAGCA 1823  
Qy 178 LysSerLeuProHisAspGluLeuGlyGlnValThrAlaAspAspPheAlaLeuGlu 198  
Db 1824 CCAGGGCGTGGGCACGACGCGCGCGCGCAACAC-----CTCCA 1865  
Qy 198 uAspCysIleLeuAsnGlyAspPheSerIleCysGluAspVal-----ProAlaGlyAs 216  
Db 1866 GGAAGCCGTG-----CGCGACCTCGGCCACCGTGCCAAACGA 1901  
Qy 216 pProAlaGlyArgLeuValAsnProThrAlaAlaPheAlaIleAspIleSerGlyProAl 236  
Db 1902 CGAGTCCGGCAAGAGCAACGAGGCAAGCGCGCATCGCC-----TTGAGCGGTCCGGC 1955  
Qy 236 aPheSerAlaThrThrIleProProValProThrLeuSerSerProGluLeuAlaAlaG 256  
Db 1956 AGGATCCGCGCGCGACGCGCGGAGCTGACCTGGCGCGCGCGAGCAGCGTCGACAG 2015  
Qy 256 nLeuAlaGluLeuTyrrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrrGly 276  
Db 2016 CGTGGCGCGCGAGAACACAGAGGTGACCGCGCGGAGAGAGGTGTGTATCAACGCGCGC 2075  
Qy 276 rAspGluIleThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLe 296  
Db 2076 CGAC-----ATCGCGCTGTTCCGCCAGGGCGGC----- 2103  
Qy 296 uAspAlaValSerIleGlySerAspGlyThrValAspProPheSerGlnLeuPheArgAl 316  
Db 2104 -----GAGCTGCGCCAGAT 2117  
Qy 316 aThrPheValGlyValGluThrGlyProPheValSerGlnLeuLeuValAsnSerPheTh 336  
Db 2118 CACC-----CACCAGGCGCGCGATGCTGCTGCGAGCGCGCAGAGAACACGACATCCG 2165  
Qy 336 rIleAspAlaIleThrValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrrMe 356  
Db 2166 CTTGGAAGCC-----GAGCAGAGCGTGGAGTCTAGCGCGCCAGCAGCAGCGTGTCT 2216  
Qy 356 tValAspPheAspGluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluG 376  
Db 2217 GGTCAACCGCAAGGAGCACATCACCTGATGTGGCGCGC----- 2256  
Qy 376 uLeuAspGluLeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPheVa 396  
Db 2256 ----- 2256  
Qy 396 lAspAsnIleAsnThrGluAlaTyrr-----ArgGlySerLeuIleLeuLeuGluLe 413  
Db 2257 -----GCCTACTGACCTCAAGGCGGCAACATC-----GAGCT 2291  
Qy 413 uGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSerAspArgGlnAlaG 433  
Db 2292 GGGC-----ATGCCCGGC-----AA 2306  
Qy 433 yPheValAsnPheGlyThrSerHisTyrrPheArgLeuIleGlyAlaAlaGluLeuAlaG 453  
Db 2307 CTTTGTGTCAAGCGCGGAGGACACATGTCAT-----GTGCGCGCGCGCAGCTGGAAGC 2360

Qy 453 nArgAlaSerCysTyrrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeuGlu 473  
Db 2361 CGAGTCCCGCAGTTTCGAGGTGGCGAGACGCGCGCG--GTTCTCTCAAGCAACTGG 2419  
Qy 473 Y-Gly-----ThrLeuHisAsn-----ThrI 480  
Db 2420 ACGGGCAGCGGCATGCCCAACGTCCTATACCATCATGCCAATGCCAGCGGAGGTCA 2479  
Qy 480 leAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuL 500  
Db 2480 TCGAGGGCGTCAACGAGCGCGGAGCGCTACCCAGCTGCTGCAGAAAGATCGCATGAACA 2539  
Qy 500 ysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrrLeu---- 518  
Db 2540 TCGCAAGGTTCGATATGAAACACACAGAAATCGCCGCTCCGCAATGTCGCGGATAGCAG 2599  
Qy 519 -----LeuProGlnAlaIleGlnValGlySer-----ProThrHisProSerTyrrProSerG 536  
Db 2600 CGGCGGTGGCGCGCGGTTGCAGTCGCAAGTCTCTGAGTGGCGCCGACGCGCGAGCGCG 2659  
Qy 536 lyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspA 556  
Db 2660 GAGTGTCTCTCAGCGAAGGGAA-----ATCAGCCTGGCGA 2695  
Qy 556 rgGlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeuLeuA 576  
Db 2696 AAGGC-----GTATTGCGGATTCATCGACTATTTCACGGTCC 2734  
Qy 576 snPhe-----GluGlyAlaCysLeuThrTyrrGluGlyGluIleAsnLysLeuAlaValAsn- 594  
Db 2735 GCCTGCGCGACGAGGACTATGTCCTGGCAGGCAAGGACTACGTGATGCGCGCAACG 2794  
Qy 595 -----ValAlaPheGlyArgGlnMetLeuGlyIle-----HisTyrrArgPheAspGlyIle- 611  
Db 2795 GCCACATCTACTTCGCGCAGAGGTTCGCGCGCTGGCGGAGCTGGAGGCTGGAGAGCCTGC 2854  
Qy 612 -----GlnGlyLeuLeuLeu-----GlyGlu 619  
Db 2855 AGCGCCAGCGTCTGTTTCATTACGMAATGACCCATGCTGCGCAGCACCAGCATGGCGTCA 2914  
Qy 619 hrIleThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluGluAlaThrPheG 639  
Db 2915 ACGTCTCTCTGTGGGGGCGCTACCCAGCAAGCCAGGAGTTTCTCTAGGCGCACAGTACG 2974  
Qy 639 luPheArgLeuPheThrGlyGluValIleValLeuPheGlnAspGlyThrPheSerIleA 659  
Db 2975 CTTACCGGTTGGAGCGCGGAAAGACGTTGAAG-----GACTACAACATCGAACAGC 3025  
Qy 659 spGlyAspMet 662  
Db 3026 AGGAGACATC 3036

RESULT 15  
US-09-266-965-76/c  
; Sequence 76, Application US/09266965  
; Patent No. 6495348  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D  
; APPLICANT: Mao, Y  
; APPLICANT: Varoglu, M  
; APPLICANT: He, M  
; APPLICANT: Sheldon, P  
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster  
; FILE REFERENCE: 600.456US1  
; CURRENT APPLICATION NUMBER: US/09/266,965  
; CURRENT FILING DATE: 1999-03-12  
; EARLIER APPLICATION NUMBER: US 08/624,447  
; EARLIER FILING DATE: 1996-08-19  
; EARLIER APPLICATION NUMBER: PCT/US94/11279  
; EARLIER FILING DATE: 1994-10-06  
; EARLIER APPLICATION NUMBER: US 08/133,963  
; EARLIER FILING DATE: 1993-10-07

; NUMBER OF SEQ ID NOS: 145  
; SOFTWARE: Fast-Seq for Windows Version 3.0  
; SEQ ID NO 76  
; LENGTH: 53500  
; TYPE: DNA  
; ORGANISM: Streptomyces lavendulae  
US-09-266-965-76

Alignment Scores:  
Pred. No.: 2.46 Length: 53500  
Score: 117.00 Matches: 151  
Percent Similarity: 32.48% Conservative: 64  
Best Local Similarity: 22.81% Mismatches: 207  
Query Match: 3.32% Indels: 240  
DB: 4 Gaps: 36

US-10-691-383-2 (1-676) x US-09-266-965-76 (1-53500)

```
Qy 2 LeuCyHisAlaAlaAspThrThrArgGlySerProMetProAspThrGlyValLeuArg 21
Db 17507 CTGGGACACGCTCTCGATGCTACCTCGAGGACGAGATCACCTCGACGAGCTGTGGCGCG 17448

Qy 22 -----LeuLeuThrSerGluInArgAlaLysGlyTrpArgArgIn----- 35
Db 17447 TCCACACCTACTCCCGCGCCGATCCGCGAGATGCTCAGCGAGCGCGAAGCGCGACTACCT 17388

Qy 36 -----LeuGlu-GlyGluLysSerLeuGlyPheHisProSerGluThrPr 50
Db 17387 CCAGGGGCGCATGACTGAACAGCGGACCGGTACGAGCGGACACCCCTCTCGGGCGC-- 17330

Qy 50 oTyriLeuLysThrLeuGluGlySerGluThrTrpLys-LysValLysLeuPro---ThrA 69
Db 17329 -----CGTGGACGGTGGCGTCCGCGCGGATACCGCTCGCCCGGACCG 17286

Qy 69 spGlyLeSerAlaSerLysLeuGlyLysLeuMetAlaArgValArgIleAlaThrA 89
Db 17285 AT-----GTGAACCTCCCTGGAGCAGGTGCTGACGCGCGCG 17247

Qy 89 laLeu-----AlaValLysLeuAlaAlaPro----- 97
Db 17246 CACAGCTCAAGCGCGGAGCGGTGGAGCGGCGGTGGCCCTGCGCGCGCCCAACAGCC 17187

Qy 98 -----CysLeuAlaPheAspGluValThrAlaSerGlyVal----- 109
Db 17186 CGCGGTGGTTCGTCCTCTCTCTCTGTCGATCCGCGCGCAGCCCTGCTGCTGCTG 17127

Qy 110 -----PheProGluLysLys----- 115
Db 17126 AACCCGACACCCCGCGCGGACCGCGCGGTGCTGCGGGTGGCGGGGCGGCGAGGT 17067

Qy 116 -----HisThrGlyGluGlyArgHisLeuGlnThrCysThrAsnSerAspA 131
Db 17066 CCTGTGTCGTCGCGGACCGCGGCGGCTCGGCTG-----ACGTGACCGGCTCGCCCG 17010

Qy 131 sp-----AlaLeuAspProThrAlaProAsnArgArgAspA 143
Db 17009 GAGAACCTCCGGCGCCCGCCCGCGTGTCTGCTCCCGACCTCGGGGTGCGACCGTGCGA 16950

Qy 143 snValAlaPheAlaSerArgArgAspAla-----AlaArgArgGluArgAspG 159
Db 16949 GCAAGCTCGTCGCGCGCGAGGAGAGACCTGCTCGCGGAGGCGCGCTACCGCGACG 16890

Qy 159 lyThrGlyThrValCysGlnIleThrAsnGlyGluThrAspLeu-----AlaThrMetP 177
Db 16889 GGGTCGGG-----CTGACGGGAGAGGACACCTGCTGCTCGCGGTGCGCGTGT 16842

Qy 177 heHis-----Lys 180
Db 16841 CCACGCGTACGCGCTGGGTGGTGTTCGGCGGACTGCTGACGGGTGCGCGGTGGCGCC 16782

Qy 180 erLeuProHisAspGluLeuGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspC 200
Db 16781 CCGTACCGCGGACCGCGCTCGGCGCGCATCGCGCGGAG----- 16744
```

```
Qy 200 ysIleLeuAsnGlyAspPheSerIleCysGluAspValProAla----- 214
Db 16743 -----CTGTCGGTGTGGCGACCGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 16689

Qy 215 -----GlyAspProAlaGlyArgLeu---ValAsnProThrAlaAlaP 228
Db 16688 CGACCGCGGCGTGGCGGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCTCCCGGTCTCCGCG 16629

Qy 228 heAlaIleAspIleSerGlyProAlaPheSerAlaThrThrIleProProValProThrL 248
Db 16628 TGGCCATGTTGGTGGTGGCGCGCGGTG----- 16603

Qy 248 euSerSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAlaArgA 268
Db 16602 -----GACGAGCAGCTGGACCGCGCTTACCGAGGCGGTTCGGGACCGGTCTCGCCGCA 16548

Qy 268 spValProPheMetGlnTyrGlyThrAspGluIleThrThrAlaAlaAlaAsnLeuA 288
Db 16547 AC-----TACGGT-----TCCACGAGACCGGCGCGGTCTCG 16515

Qy 288 laGlyMetGlyGly-----PheProAsnLeuAspAlaV 299
Db 16514 CCGACCGCGCGCGGTGGAGCCCTTGGCGCGGTCTCCCTCGCGGGGTGGAGTGGC 16455

Qy 299 alSerIleGlySerAspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheV 319
Db 16454 AACTGACCGCGCGGAGGCGGTGTCGCGCC----- 16423

Qy 319 alGlyValGluThrGlyProPheValSerGlnLeuValAsnSerPheThrIleAspA 339
Db 16422 -----GCCGCGACCCCGCGGCG 16407

Qy 339 laIleThrValGluProLysGlnGluThrPheAla---ProAspLeuAsnTyrMetValA 358
Db 16406 TGCTGACGTACCGGTGCGACCGCGCGGTACGCCATGGCGCATCTCCGCGTGGCGCTGC 16347

Qy 358 spPheAsp-----GluTrpLeuAsnIleGlnAsnGlyGlyProProA 372
Db 16346 CCGGGGCGCTCGCATCTCTGGGACGCGAGACCGCGCGATCCCGCGGGCGGCGCTGGG 16287

Qy 372 laGlyProGluLeuLeuAspGluLeuArgPheIleArgAsnAlaArgAspLeuAlaA 392
Db 16286 TCTCCCGCTGGAGATCGAGAGGTGCTGGCGGTATCCCGGAC----- 16243

Qy 392 rgValSerPheValAspAsnIleAsnThrGluAlaTyrArgGly-----S 407
Db 16242 -----GTGGTGAATGTGGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGG 16194

Qy 407 erLeuIleLeuLeuGluLeuGlyValaPheSerArgProGlyIleAsn----- 422
Db 16193 ACGGATCTGTCGCGAGGTCTCGCG-----CGCGGCGCGCGGCTCACCCCGAGCGCTGC 16137

Qy 423 -----GlyProPheIleAspSerAsp----- 429
Db 16136 GCGACGACCGCGCGGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 16077

Qy 430 -----ArgGlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgL 445
Db 16076 AGAGCTTGGCGTCAACGCGCGCGCGCAAGGTGGCG-----GCCGCGCTCGCTTACCGCC 16023

Qy 445 euIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnVal---- 463
Db 16022 TCACCGGAGCGCG-----GCCGAGCGCGCGCGCGGTACAGGCATCCGAAGTGTCT 15969

Qy 464 -----HisArgPheAlaArgProGluAlaLeuGly-----GlyThrLeuHisAsnThrI 480
Db 15968 TCGCGCTTCACGACTTGGCGCGCTTGGAGGCACTCGCCAGGCTGGCGGACGCGCTCTCC 15909

Qy 480 leaIaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuL 500
Db 15908 TCGCCGCGGAGCTGGGTGCGACCGCGATCCCTG-----GAGTGGCTGC 15864
```

```
Qy      500 ysArgValAlaGlu-----IleAsnAlaAlaGlnAsnProAsnAsnGluV 515
Db      15863 TGGCGACGGCCACCGCTCTGGGGGTGTGACCAACCGGGGGCAAGAGCCCGGGGACCGGG 15804
Qy      515 al 515
Db      15803 TC 15802
```

Search completed: September 19, 2004, 10:15:25  
Job time: 3134.41 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 19, 2004, 03:56:03 ; Search time 869.364 Seconds  
(without alignments)  
3303.314 Million cell updates/sec

Title: US-10-691-383-2  
Perfect score: 3528  
Sequence: 1 MLCHAADTRGSPMPPTGV.....SIDGDMCSGLVVTGVADCA 676

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgr2\_1/USPTO\_spool/US10691383/runat\_17092004\_102655\_1733/app\_query.fasta\_1.1230  
-DB=N\_Geneseq\_29Jan04 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blos62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10691383 @CGN 1.1 653 @runat\_17092004\_102655\_1733 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq\_29Jan04.\*  
1: Geneseq1980s.\*  
2: Geneseq1990s.\*  
3: Geneseq2000s.\*  
4: Geneseq2001as.\*  
5: Geneseq2001bs.\*  
6: Geneseq2002s.\*  
7: Geneseq2003as.\*  
8: Geneseq2003bs.\*  
9: Geneseq2003cs.\*  
10: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3528	100.0	2931	3	AA10305 Fucus van
2	3528	100.0	2931	6	Aad28155 Fucus dis
3	626	17.7	1797	4	AAD11050 Corallina
4	623	17.7	1794	2	AAV56020 Haloperox
5	614.5	17.4	1791	2	AAV56021 Haloperox
6	131	3.7	9579	7	AAD54222 Streptomy
7	131	3.7	52101	7	AAD54217 Streptomy
8	129.5	3.7	11679	8	ADA30090 DNA encod

9	124	3.5	18930	7	ACA43739 Prokaryot
10	123.5	3.5	68750	3	Aaz55887 Sorangium
11	123	3.5	1962	7	ACF70799 Photorhab
12	123	3.5	5820	6	ABs78676 A. orient
13	123	3.5	110000	7	Continuation (39 o
14	123	3.5	110000	7	Continuation (10 o
15	118	3.3	3426	4	Aaf61025 P. putida
16	118	3.3	349980	6	ABQ81847 Bifidobac
17	117.5	3.3	3505	7	ACA35977 Prokaryot
18	117.5	3.3	110000	4	Continuation (24 o
19	117	3.3	3456	2	Aat85397 Pseudomon
20	117	3.3	53500	3	Aac55842 Complete
21	117	3.3	53500	9	Ade10261 S. lavend
22	116.5	3.3	3971	7	AAD47845 Haemophil
23	116.5	3.3	11883	3	AAA10263 Bordetell
24	116.5	3.3	12036	2	AAQ04668 FHA struc
25	116.5	3.3	19547	4	AAS59601 Propionib
26	116.5	3.3	19547	7	ACF64530 Propionib
27	116.5	3.3	30121	2	ACF42063_18
28	116.5	3.3	110000	2	Continuation (18 o
29	115.5	3.3	2271	2	AAQ06844 Amylase g
30	115	3.3	65140	4	AAD17184 Streptomy
31	115	3.3	125401	4	AAD17186 Streptomy
32	114	3.2	47988	9	ADE86070 Streptomy
33	114	3.2	349980	6	ABQ81845 Bifidobac
34	113.5	3.2	2824	6	ABN89350 Chicken N
35	113.5	3.2	2824	6	ABN89351 Chicken N
36	113.5	3.2	10419	3	AAAS5188 Cenarchae
37	113.5	3.2	42432	3	AAAS5187 Cenarchae
38	113	3.2	88421	6	AAL40781 8842Int g
39	112	3.2	2520	2	AAQ40580 Glutaryl-
40	111.5	3.2	75236	7	ABV75557 Saccharop
41	111	3.1	25801	4	AAD11177 Pseudomon
42	110.5	3.1	4172	4	ABL13509 Drosophil
43	110.5	3.1	110000	6	ABQ67196_3
44	110.5	3.1	110000	6	Continuation (4 of
45	110.5	3.1	110000	6	Continuation (5 of
c	110.5	3.1	110000	6	ABQ69245_28

ALIGNMENTS

RESULT 1	AAAA10305 standard; cDNA; 2931 BP.
AAAA10305	
ID	AAAA10305
AC	AAAA10305;
XX	
XX	
DF	03-JUL-2000 (first entry)
XX	
XX	Fucus vanadium bromoperoxidase cDNA.
DE	
XX	Vanadium bromoperoxidase; vanadium haloperoxidase; alga; ODA;
KW	o-dianisidine oxidation; vanadium ion; halogenation; antimicrobial agent;
KW	phenolic adhesive production; ss.
OS	Fucus sp.
XX	
XX	Key Location/Qualifiers
PH	228..2258
FT	/*tag= a
FT	/product= "Fucus vanadium bromoperoxidase"
XX	
XX	WO200015771-A1.
PD	23-MAR-2000.
XX	
XX	27-AUG-1999; 99WO-US019893.
XX	
PR	10-SEP-1998; 98US-00151189.
XX	
PA	(REGC ) UNIV CALIFORNIA.
XX	
PI	Vreeland V, Ng KL;



Qy 501 ArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuPro 520  
 Db 1728 CGTGGCGGAGATAAATCGCGCGAGAAATCCCAACAGAGGTACCTACTCTTCCA 1787  
 Qy 521 GlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrCln 540  
 Db 1788 CAAGCTATCCAAAGTGGGATCGCCAAACGACCCCTTCTACCGTCCGCGCACGCTACCACAA 1847  
 Qy 541 AsnGlyAlaPheAlaThrValLeuLeuAlaLeuIleGlyLeuAspArgGlyGlyGluCys 560  
 Db 1848 AATGGAGCATTTGCCACAGTTCTGAAGGCCCTCATTTGGCCTAGATCGGGAGGTGAGTGC 1907  
 Qy 561 PheProAsnProValPheProSerAspGlyLeuGluLeuIleAsnPheGluGlyAla 580  
 Db 1908 TTCCCTTAACCCCGTGTTCCTCAAGCGATCAGCGCCTGGAACCTAATCAACTTCGAAGGGGCA 1967  
 Qy 581 CysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGln 600  
 Db 1968 TGCCTTACATATGAGGAGAGATCAACAAGCTCGCGTCAACGTTCGCTATTTGGGAGGCAG 2027  
 Qy 601 MetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuLeuGlyGluThrIle 620  
 Db 2028 ATGCTGGGCATCCACTATCGTTCGAGGTATCCAGGCCCTACTTCTCGAGAGACAAATC 2087  
 Qy 621 ThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluGluAlaThrPheGluPhe 640  
 Db 2088 ACTGTACGAACACTTCACAGGAGCTGATGACGTTCCGCGAGGAAGCCACCTTTGAATTC 2147  
 Qy 641 ArgLeuPheThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIleAspGly 660  
 Db 2148 CGCTATTATCCAGGAGGATCATCAAACTTTCCAGACCGGACATTTCTCCATCGATGGA 2207  
 Qy 661 AspMetCysSerGlyLeuValTyrThrGlyValAlaAspCysGlnAla 676  
 Db 2208 GATATGTGTTCCGGTTTGGTTTACACTGGCGTGGCGACTGCCAGGCT 2255

RESULT 2  
 AAD28155  
 ID AAD28155 standard; cDNA; 2931 BP.  
 XX  
 AC AAD28155;  
 XX  
 DT 07-AUG-2003 (revised)  
 DT 01-MAY-2002 (first entry)  
 XX  
 DE Fucus distiches vanadium haloperoxidase encoding cDNA.  
 XX  
 KW Vanadium haloperoxidase; vanadium bromoperoxidase; o-dianisidine; ODA;  
 KW epoxide; antimicrobial agent; phenolic adhesive; halohydrin;  
 KW industrial catalysis; enzyme; catalyst; ss.  
 XX  
 OS Fucus distichus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 228..2258  
 FT /\*tag= a  
 FT /product= "vanadium haloperoxidase"  
 XX  
 PN WO200200838-A2.  
 XX  
 XX 03-JAN-2002.  
 XX  
 XX 07-JUN-2001; 2001WO-US018602.  
 XX  
 XX 19-JUN-2000; 2000US-00596794.  
 XX  
 XX (REGC ) UNIV CALIFORNIA.  
 XX  
 XX Vreeland V;  
 XX  
 XX WPI; 2002-154651/20.  
 XX P-PSDB; AAE17501.  
 XX

PT isolated nucleic acid for, e.g. to halogenate various substrates  
 PT particularly vanadium bromoperoxidase polypeptide which consists of a  
 PT haloperoxidase polypeptide with catalytic helical frame.  
 XX  
 XX Claim 4; Page 49-53; 56pp; English.  
 XX  
 CC The invention relates to nucleic acid encoding vanadium haloperoxidase  
 CC particularly vanadium bromoperoxidase polypeptide which consists of a  
 CC catalytic helical frame that complexes a vanadium ion and catalyses the  
 CC oxidation of o-dianisidine (ODA). The invention also provides  
 CC recombinantly produced vanadium haloperoxidases. The polypeptides of the  
 CC invention can be fused to other proteins to allow quantification or  
 CC localisation of the linked protein. They can be used to halogenate  
 CC various substrates including proteins; to produce epoxides from alkenes,  
 CC halogenated ketones from alkyne; to produce alpha, gamma-halohydrins  
 CC from cyclopropanes and to produce dihalogenated products from alkenes and  
 CC alkynes. They can also be used in signal generating systems in place of  
 CC horseradish peroxidase; as a component in assays; as enzymatic  
 CC antimicrobial agents; and in the production of phenolic adhesives. The  
 CC polypeptides can be used in industrial catalysis in a variety of  
 CC contexts, e.g. catalyst for halogenation, oxidation, and epoxidation  
 CC reactions. The present sequence is Fucus distiches vanadium  
 CC haloperoxidase encoding cDNA. (Updated on 07-AUG-2003 to correct OS  
 CC field.)

XX SQ Sequence 2931 BP; 694 A; 776 C; 784 G; 677 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.:	0	Length:	2931
Score:	3528.00	Matches:	676
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-691-383-2 (1-676) x AAD28155 (1-2931)

Qy	1	MetLeuCysHisAlaAlaAspThrThrArgGlySerProMetProAspThrGlyValLeu	20
Db	228	ATGCTTTGCCATCGAGCGGACACGACAAGAGGCTCTCTATGCTGACACCGGAGTGCTT	287
Qy	21	ArgLeuLeuThrSerGluGlnArgAlaLysGlyTrpArgArgGlnLeuGluGlyLys	40
Db	288	CGGTGCTCATCAGACGACGCGCTAAAGTTGGAGAGCCAGCTTAGAGGGGGAGAAA	347
Qy	41	SerLeuGlyPheHisProSerGluThrProTyrIleLysTyrLeuGluGlySerGluThr	60
Db	348	TCACTAGGTTTTCATCCAGCGAGACGCCCTTATATCAAGTACTTTGGAGGGCTCTGAGACT	407
Qy	61	TrpLysLysValLysLeuProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIle	80
Db	408	TGGAAGAAGGTTAAGCTTCCAACGGACGCATATCGGCTTCCAAGATCTCGGTAAAAATT	467
Qy	81	MetAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAla	100
Db	468	ATGGCCAGGGTCCGCATCGCTACCGCTTGGTGTGTTACTTGGCCGCCACCTGTTTGCA	527
Qy	101	PheAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGlyGlyGly	120
Db	528	TTCCAGGAGGTACAGCCAGTGGTGTGTTTCCCTGAGGAACACAGCACCCCGGGAGGGA	587
Qy	121	ArgHisLeuGlnThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArg	140
Db	588	AGACACCTCCAGACCTGTACAAACTCCGACGATCGCGCTGGATCCGACGGGCCGAATAGA	647
Qy	141	ArgAspAsnValAlaPheAlaSerArgArgAspAlaAlaArgGluAlaArgAspGlyThr	160
Db	648	AGGGACAACGTAGCTTTTGGCTCGCGCGCGATGCCGCCAGGCGAGAACGTGACGGGACA	707
Qy	161	GlyThrValCysGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSer	180
Db	708	GGGACTGTCTGCCAAATCACTAAACGGAGAACTGATTGGCTTACATGTTCCCAAGTCT	767

```
QY 181 LeuProHisAspGluLeuGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCys 200
DB 768 CTGCCACACGATGACCTGGACAGGTAAACCGCAGACGACTTCGCTATCTCTCGAGGACTGC 827
QY 201 IleLeuAsnGlyAspPheSerIleCysGluAspValProAlaGlyAspProAlaGlyArg 220
DB 828 ATCTTAAACGAGATTTTCAGCATTTGCGAGACGTGCTCGCGGAGACCCCGGGGTGCG 887
QY 221 LeuValAsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThr 240
DB 888 CTGGTCAATCTTACCGCTGGCTTGGCCATGACATATCCGGTCCCGCATTTCTCGGTACG 947
QY 241 ThrIleProProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeu 260
DB 948 ACAATACCCCGGTACCTACCTTCTCTCTGAGCTCGCGCTCAGTTGGCGGAGCTA 1007
QY 261 TyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGluIleThr 280
DB 1008 TACTGGATGGCGCTGGCCAGGATGTACCTTTATGCAGTATGGCACCGAGAAATTACC 1067
QY 281 ThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSer 300
DB 1068 ACTACCGCGGACGACCACTCGCTGGAAATGGAGGCTTCCAAATCTGGACGCGGTGCG 1127
QY 301 IleGlySerAspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGly 320
DB 1128 ATAGGTCCTGATGATGCTGGAGCCCGTTCTCCAGCTCTTCGAGGCGACTTCGTGGT 1187
QY 321 ValGluThrGlyProPheValSerGlnLeuValAsnSerPheThrIleAspAlaIle 340
DB 1188 GTTGAAACGGGGCCCTTGTCTCTCAGCTGCTCGTGAACAGCTTCACCATCGACGCTATT 1247
QY 341 ThrValGluProIysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPheAsp 360
DB 1248 ACGGTGCAACCGAAGCAGGAGACATTCGCCCGCGACTTGAACATATATGGTCGATTTGAC 1307
QY 361 GluTrpLeuAsnIleGlnAsnGlyGlyProAlaGlyProGluGluLeuAspGluGlu 380
DB 1308 GAATGGCTGAACATTCAGATGTGGACCCCGCGCGCCCGCCGAGATTAGACGAAGAG 1367
QY 381 LeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsn 400
DB 1368 CTGGCTTTATCGTAACGCCCGCGACCTGGCCAGGCTCTCTCGTGGCAATATCAAC 1427
QY 401 ThrGluAlaTyrArgGlySerLeuIleLeuGluLeuGlyAlaPheSerArgProGly 420
DB 1428 ACCGAAGCTTATCGCGGTCTCTTATCTTACTTGTAGCTGGGAGCTTTCAGCAGCGCCGT 1487
QY 421 IleAsnGlyProPheIleAspSerAspArgGlnAlaGlyPheValAsnPheGlyThrSer 440
DB 1488 ATCAACGGTCCATTCATCGACAGTGTATCGGCGGCGGCTTCGTCAACTTCGGCAGCTCT 1547
QY 441 HisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLys 460
DB 1548 CACTACTTCAGATTGATAGTGGCGCGAGCTGGCGCAGGCTGCTCGTGTACCAAAAG 1607
QY 461 TrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThrIleHisAsnThrIle 480
DB 1608 TGGCAGGTGATCATGTTGACGCGCCGAGGCTCTCGGGGGTACCCCTCCACACACCATC 1667
QY 481 AlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLys 500
DB 1668 CGGGGGATCTAGATGACAGACTTCGACATCTCCCTCTTGGAAATGATGAGCTTTGAAA 1727
QY 501 ArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuPro 520
DB 1728 CGTGTGGCGAGATAAATCGGGCGAGAAATCCCAACACGAGGTACCTACCTCTTCTCCA 1787
QY 521 GlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGln 540
DB 1788 CAGCTATCAAGTGGATCGCCAACGACCTTCTTACCGCTCCGGCCACGCTTACCACAA 1847
QY 541 AsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCys 560
```

```
DB 1848 AATGGAGCATTTGCCACAGTTCTGAAGCCCTCATTTGGCCTAGATCGGGAGGTGAGTGC 1907
QY 561 PheProAsnProValPheProSerAspAspGlyIleuGluLeuIleAsnPheGluGlyAla 580
DB 1908 TTCCTTAACCCCGTGTTCCTCAAGCGATGACGGCTCGGAACATACTCACTTCGAAGGGGCA 1967
QY 581 CysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGln 600
DB 1968 TGCCTTACATATGAGGAGAGATCAACAAGCTCGCGTCAACGTCCGATTTGGAGGCGAG 2027
QY 601 MetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThrIle 620
DB 2028 ATGCTGGGCATCCACTATCGTTTCGACGATATCCAAGGCTACTTCTCGAGAGACAATC 2087
QY 621 ThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluAlaThrPheGluPhe 640
DB 2088 ACTGTACGAACACTTCACCAAGGAGCTGATGACGTTCCGCCGAGGAAGCCACTTTGAATTC 2147
QY 641 ArgLeuPheThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIleAspGly 660
DB 2148 CGCTTATTCACCGAGAGGTCACTCAAACTTTTCAGGACGGGACATTCCTCATCGATGGA 2207
QY 661 AspMetCysSerGlyLeuValTyrThrGlyValAlaAspCysGlnAla 676
DB 2208 GATATGTGTTCCGGTTTGGTTTACACTGGCGTGGGAGCTGCCAGGCT 2255

RESULT 3
AAD11050
ID AAD11050 standard; cdna; 1797 BP.
XX
AC AAD11050;
XX
DT 24-SEP-2001 (first entry)
XX
DE Corallina officinalis vanadium bromoperoxidase (cVBPo) cDNA.
XX
KW Marine red algae; vanadium bromoperoxidase; VBPo; haloperoxidase;
KW signal generation; analytical system; anti-microbial agent;
KW industrial application; biological material; avidin; metabolite;
KW lipopolysaccharide; human chorionic gonadotropin; luteinising hormone;
KW vitamin; intact cell; drug; narcotic; ss.
XX
OS Corallina officinalis.
XX
FH Key Location/Qualifiers
CDS 1..1797
FT /*tag= a
FT /product= "Corallina officinalis vanadium bromoperoxidase
FT (VBPo)"
FT primer_bind complement(1..21)
FT /*tag= b
FT /bound_moiety= "PCR forward primer (AAD11054)"
FT primer_bind 1150..1170
FT /*tag= c
FT /bound_moiety= "PCR reverse primer (AAD11055)"
FT misc_binding complement(1420..1449)
FT /*tag= d
FT /bound_moiety= "PCR probe (AAD11053)"
FT primer_bind complement(1420..1442)
FT /*tag= e
FT /bound_moiety= "PCR upstream primer (AAD11051)"
FT primer_bind 1642..1664
FT /*tag= f
FT /bound_moiety= "PCR downstream primer (AAD11052)"
XX
XX WO200153494-A2.
XX
XX 26-JUL-2001.
XX
XX 24-JAN-2001; 2001WO-US002383.
XX
XX 24-JAN-2000; 2000US-0177764P.
XX
XX
```



XX (REGC ) UNIV CALIFORNIA.  
 XX Butler A, Carter JN;  
 XX WPI: 2001-457611/49.  
 XX P-PSDB; AAE05750.  
 XX Novel vanadium haloperoxidase, especially vanadium bromoperoxidase,  
 PT obtained from *Corallina officinalis* for use in signal generation in  
 PT analytical systems.  
 XX Claim 25; Fig 1; 68pp; English.  
 XX The present sequence is *Corallina officinalis* (marine red algae) vanadium  
 CC bromoperoxidase (cVBP) cDNA. The vanadium haloperoxidase, especially  
 CC vanadium bromoperoxidase (VBPO) is useful for signal generation in  
 CC analytical systems; and as anti-microbial agents. The VBPO is also useful  
 CC for industrial applications. The VBPO is also useful for detecting a wide  
 CC variety of chemical and biological materials such as amino acids,  
 CC peptides, polypeptides, proteins including enzymes, avidin, antibodies  
 CC and antigenic proteins, carbohydrates including monosaccharides,  
 CC polysaccharides and lipopolysaccharides, hormones such as human chorionic  
 CC gonadotropin, thyroid stimulating hormone, luteinising hormone,  
 CC thyroxine, follicle stimulating hormone, parathyroid hormone and growth  
 CC hormone, metabolites such as glucose, lactate and pyruvate,  
 CC oligonucleotides, nucleic acids, vitamins such as B12 and biotin, intact  
 CC cells from various organisms including microorganisms and drugs such as  
 CC narcotics, therapeutic and those abused  
 XX SQ Sequence 1797 BP; 394 A; 480 C; 513 G; 410 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 6,22e-49 Length: 1797  
 Score: 626.00 Matches: 176  
 Percent Similarity: 45.00% Conservativeness: 76  
 Best Local Similarity: 31.43% Mismatches: 220  
 Query Match: 17.74% Indels: 88  
 DB: 4 Gaps: 15

US-10-691-383-2 (1-676) x AAD11050 (1-1797)

Qy 168 AsnGlyGluThrAspLeuAlaThrMetPheHisLysSerLeuProHisAspGluLeuGly 187  
 Db 148 AACGGAGATCCAGCTTTCATGCTTACAAAGGGTCTTCGCGCATGACGCAATGCG 207  
 Qy 188 Gln---ValThrAlaAspPheAlaLeuGluAspCysIleLeuAsnGlyAspPhe 206  
 Db 208 GCTATTATGACCCGACGACTTCTTGGCTTCGTCGTCGATCAATAGTGGAGATGAG 267  
 Qy 207 SerileCysGluAspVal-----ProAlaGlyAspProAlaGlyArgLeu----- 221  
 Db 268 AAGGAGATCCGACCTCACATTGGGGCCAGCTCGCGACCCGACACTGGCTTACCAATC 327  
 Qy 222 -----ValAsnProThr 225  
 Db 328 TGGCGCTCGGATTGGCGAAFTCTCTCGAGCTCGAAGTGGCAGGATGGGAAACAGCTCT 387  
 Qy 226 AlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThrIleProProVal 245  
 Db 388 GCCGGCTCCTACCTTCGACCTTGAGGGCCCGACCGCGCAGTCGATGTCGACCCGCG 447  
 Qy 246 ProThrLeuSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyrrpMetAlaLeu 265  
 Db 448 CTTGTGCTCAGAGCCCTGAGCTCATCGCGAGATAGCAGACTGTACCTGATGGCGCTT 507  
 Qy 266 AlaArgAspValProPhe-----MetGlnTyrr 274  
 Db 508 GGACCGGAAATCGAGTTTAGCGAGTTTTCATTTCCCAAGAACGACGACGATATATTCAGTTT 567  
 Qy 275 GlyThrAspGluIleThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPhePro 294  
 Db 568 GCTATTATGAGCTTAACGGGCTAGATGGTTCAACACGCGCGCAATGCTCGGAGATCCG 627

Qy 295 AsnLeuAspAlaValSerIleGlySerAspGlyThrValAspProPheSerGlnLeuPhe 314  
 Db 628 CCTCGGAAATCCGTCGCGCGTGGGTGAGTGACTGTC-----GGAAACTTGTTC 678  
 Qy 315 ArgAlaThrPheValGlyValGluThrGlyProPheValSerGlnLeuVal----- 332  
 Db 679 CGCGGTATTCTTCAGGCTCTGAGGTCGGGCCGCTATCTCAGCCAGTACATCATCGTTGGT 738  
 Qy 332 ----- 332  
 Db 739 AGCAAGCAGATTGGCTCGCGACGGGTGGCAAACTCTCTGAGAGCCCAATGCCGCT 798  
 Qy 333 -----AsnSerPheThrIleAspAlaIleThrValGluProLyGlnGluThr 348  
 Db 799 GATGAGTTGATGTGTAATCGCCTACGGAAGCATCACCATTAGCCAGCGGTGCGTATC 858  
 Qy 349 PheAlaProAspLeuAsnTyrrMetValAspPheAspGluTrpLeuAsnIleGlnAsnGly 368  
 Db 859 GCCACGCTCGGCGGACTTTATGACGACTTGAAGGTATTCTTGTGAGCTCCAGGAGCGCT 918  
 Qy 369 GlyProAlaGlyProGluGluLeuAspGluLeuArgPheIleArgAsnAlaArg 388  
 Db 919 GCGGACTTCGAGGCTTGGTTCGTATGACGGGAGCAGCGCTCATCCGACGATCCGC 978  
 Qy 389 AspIleAlaArgValSerPheValAspAsnIleAsnThrGluAlaTyrrArgGlySerLeu 408  
 Db 979 GATCTTCGAGCGTGGGTGCACTTTGACGCACTTGACGAGCGCTACCTCAATGCGTGCCTG 1038  
 Qy 409 IleIleLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSer 428  
 Db 1039 ATTCTGTTCGGAACCGCGCTGCTGATCCCAACATT-----CCGTTCCAGCAGAG 1092  
 Qy 429 AspArg-----GlnAlaGlyPheValAsnPheGlyThrSerHisTyrrPheArgLeu 445  
 Db 1093 GACAAGCTCGATAACACGAGCGTGTGTGNACTTCGGAGACGACACGCTGCTGAGTCTG 1152  
 Qy 446 Ile---GlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrrGlnIlyTrpGlnValHis 464  
 Db 1153 GTGACTGAGTGGCTACGCGCGCTTGAAGCGGTACGCTACGAGAAAGTTTAACTTCAT 1212  
 Qy 465 ArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHis---AsnThrIleAlaGlyAsp 483  
 Db 1213 CGTGCCTCGCGCTGAGGCTACCGGTGGTCTGATTAGCGTGAACAAATCCGACGCGAG 1272  
 Qy 484 -----LeuAspAlaAspPheAspIleSerLeuGluAsnAspGluLeuLeu 499  
 Db 1273 AAGGGCGAGAGCGTTCCTTCCTGAGTTGACTTCTGCTGTGAGAGCTTGAAGATATCCTG 1332  
 Qy 500 LysArgValAlaGluIleAsnAlaAlaGlnAsn-----ProAsnAsnGlu 514  
 Db 1333 GAGAAAGCTGAAATTAGCAATAGGAAACAGAACATTGCTGACGGAGATCTGTGACCTGAT 1392  
 Qy 515 ValThrTyrrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrrPro 534  
 Db 1393 CCTTCATTCTCTGTTCGCGAGGCAATCCCGAGGAGCGCCATTCCTATCCGCTCCACGGA 1452  
 Qy 535 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGly--- 553  
 Db 1453 AGCGGCCACGCTGTGGTGTGCTGGCGCATGTGTGACGATCTCAAGGGGCTTCTCGACTCC 1512  
 Qy 554 -----LeuAspArgGlyGlyGluCysPheProAsnProValPheProSerAspAsp 570  
 Db 1513 AACTTCCAGATCGATCAGGTGTTTCGAG-----GTCGCAAAAGATGAGGAC 1557  
 Qy 571 GlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyrrGluGlyGluIleAsnLys 590  
 Db 1558 AAGCTTGTAAAGTCGTCTTTCAAGGGAACT---CTCACTGTTCGGGTGAGCTGAACAG 1614  
 Qy 591 LeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleHisTyrrArgPheAspGly 610  
 Db 1615 CTGCGCCACAATATTGATCGCGCGGAACATGCGAGGTGTTTCACTACTTCTCTGACGAG 1674

```

Qy 611 IleGlnGlyLeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMet 630
Db 1675 TTGAGTCAATTCGCTTGGTGGACAGGTTGCGATTGGAATCTTGGAGAGCAGAGTCTG 1734
Qy 631 ThrPheAlaGluGluAlaThrPheGluPheArgLeuPheThrGlyGluValIleLysLeu 650
Db 1735 ACGTATGCGGAGAACTCTCTTCAACTTCCGGAAGTTTGTAGTGAACACTACAATCCAGATC 1794

RESULT 4
ID AAV56020
AC AAV56020;
XX
DT 09-DEC-1998 (first entry)
XX
DE Haloperoxidase, CP.BP01 encoding cDNA.
XX
KW Haloperoxidase; enzyme; CP.BP01; CP.BP02; vanadium; halide specificity;
KW bromine; iodine; ds.
XX
OS Corallina pilulifera.
XX
FH Key Location/Qualifiers
FT CDS 1..1794
FT /tag= a
FT /product= "Haloperoxidase, CP.BP01"
FT /note= "the stop codon is not indicated"
XX
XX JP10248581-A.
XX
XX 22-SEP-1998.
XX
XX 06-MAR-1997; 97JP-00070539.
XX
XX 06-MAR-1997; 97JP-00070539.
XX
XX (SAKA ) OTSUKA PHARM CO LTD.
XX
XX WPI; 1998-560733/48.
XX
XX P-PSDE; AAW80550.
XX
XX New haloperoxidase gene - and corresponding vector, transformed host cell
XX and method of preparation.
XX
XX Claim 2; Page 13-15; 20pp; Japanese.
XX
XX This cDNA encodes a haloperoxidase enzyme. The haloperoxidase genes
XX encoding the enzymes CP.BP01 and CP.BP02 are isolated from Corallina
XX pilulifera. A host cell transformed with a vector containing the
XX haloperoxidase genes can be used for the recombinant production of the
XX enzyme. The haloperoxidases of the invention are vanadium-dependent and
XX have halide specificity of bromine and iodine
XX
SQ Sequence 1794 BP; 395 A; 481 C; 505 G; 413 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.19e-48 Length: 1794
Score: 623.00 Matches: 191
Percent Similarity: 44.41% Conservative: 75
Best Local Similarity: 31.89% Mismatches: 233
Query Match: 17.66% Indels: 100
DB: 2 Gaps: 20

US-10-691-383-2 (1-676) x AAV56020 (1-1794)
Qy 141 ArgAspAsnValAlaPheAlaSerArgArgAspAlaAlaArgArgGluArgAspGlyThr 160
Db 31 CGCGCCAGGCTTCATTCGATACCGGTGTAGCTGCGCGCGAGCTTGCACCTCAAC---CGC 87
Qy 161 GlyThrValCysGlnIleThrAsnGlyGlu----- 170
Db 88 GGAGTTGTACCATCGTTTGCATAATGGGGAAGAGCTTCTCTACCGCAACCGGACCTTGAC 147

```

```

Qy 171 ---ThrAsp-----LeuAlaThrMetPheHisSerLeuProHisAspGluLeuGly 187
Db 148 AACACTGATCCGAGCTTCGCTAGCTTCACAAAGGGTCTTCGATCAGCAGCAATGGC 207
Qy 188 Gln---ValThrAlaAspAspPheAlaIleLeuGluAspCysIleLeuAsnGlyAspPhe 206
Db 208 GCTATTATCGACCCGACGACTTCTTGGCCCTTCGTCGTCGAATCAATAGTGGCGATGAA 267
Qy 207 SerIleCysGluAspVal-----ProAlaGlyAspProAlaGlyArgLeu----- 221
Db 268 AAGGAGATCGCCGACCTCACATTGGGGCAGCTCGCGACCCGAGAGCTGCTTACCAATC 327
Qy 222 -----ValAsnProThr 225
Db 328 TGGCGCTCGGATCTGGCGAATTCTTCGAGCTCAAGTGGGAGATGGGAGACAGCTCT 387
Qy 226 AlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThrThrIleProProVal 245
Db 388 GCCGGTCTCACCTTCGACCTGGAGGGCCCGACGCGCAGTCGATTGCCATGCCACCGCG 447
Qy 246 ProThrIleSerSerProGluLeuAlaIleGlnLeuAlaGluLeuThrTrpMetAlaLeu 265
Db 448 CTTGTGCTCAGAGCCCTGAGCTCGTGGCGGAGATAGCAGAGCTGTACTGATGGCGCTT 507
Qy 266 AlaArgAspValProPhe-----MetGlnTyr 274
Db 508 GGACGGCAATCGAGTTTCAGCGAGTTGATCCCAAGAACGAGAGATATTTCAGTTT 567
Qy 275 GlyThrAspGluIleThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPhePro 294
Db 568 GCTATTGATCAGCTTAACGGGCTGGAGTGGTTCAACACACCGGCAAAAGCTCGGAGATCCG 627
Qy 295 AsnLeuAspAlaValSerIleGlySerAspGlyThrValAspPropheSerGlnLeuPhe 314
Db 628 CTTGCGGAAATCGCTCGCCGCTCGCGGTGAGTGACTGTT-----GGAAACTTGTTC 678
Qy 315 ArgAlaThrPheValGlyValGluThrGlyProPheValSerGlnLeuLeu----- 331
Db 679 CGCGGTATTCTTCAGGCTCTGAGGTGCGGCCGTACCTCAGCCAGTACATCATCTGTTGGT 738
Qy 332 -----ValAsnSerPheThrIle----- 337
Db 739 AGCAAGCAGATTGCTCAGCGACAGTTGGTGAACAAACTCTCTGTGAGCCCCCAATGTGCT 798
Qy 338 -----AspAlaIleThrValGluProLysGlnGluThr 348
Db 799 GATCAGTTTGATGGTGAATTCGCTAGCGAAGCATCACCATTAGCCAGCGCTGTGCGTATC 858
Qy 349 PheAlaProAspLeuAsnTyrMetValAspPheAspGluTrpLeuAsnIleGlnAsnGly 368
Db 859 GCCACGCTGGAGCGGACTTCATGACGACTTGAAGGTATCTCTGAGCTCCAGGAGCT 918
Qy 369 GlyProProAlaGlyProGluGluLeuAspGluLeuArgPheIleArgAsnAlaArg 388
Db 919 GCGGACTTCGAGGCTTTGAGTCGTATGAGCGCGGAGACGCTCATCCGACGATCCGC 978
Qy 389 AspLeuAlaArgValSerPheValAspAsnIleAsnThrGluAlaTyrArgGlySerLeu 408
Db 979 GATCTTGCAGCGTGGGTGCACTTTGACGCACTGTACGAGCGCTTACCTCAATGGCGTCTA 1038
Qy 409 IleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSer 428
Db 1039 ATTCTGTTGGCAACGGCGTGGCTTCGATCCC-----AACCTTCGCTCCAGCAGAG 1092
Qy 429 AspArg-----GlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeu 445
Db 1093 GACAAGCTCGATAACACGAGGACGCTTTGTGAATTCGATCCGACACGCTGCTGAGTCTG 1152
Qy 446 Ile---GlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnIlyTyrTrpGlnValHis 464
Db 1153 GTGAGTGAAGTGGCTACGCGCGCTGTTGAAGCGGTACGGTACCAAGAGTTTAACTTCAT 1212

```

```

Qy 465 ArgPheAlaArgProGluAlaLeuGlyGly---ThrLeuHisAenThrIleAla----- 481
Db 1213 CGTGCCTGCGCCCTGAGGCTACCGGTCTGATTATCGTTAAACAAATCGCACCGCAG 1272
Qy 482 -----GlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeu 499
Db 1273 AAGGGCAGAGCATTTCCCTGAGGTGATCTTCTGTGTGAGAGCTTGGAGATATCTTG 1332
Qy 500 LysArgValAlaGluIleAsnAlaAlaGlnAsn-----ProAsnAsnGlu 514
Db 1333 GAGAAAGCTGAAATTAGCAATAGGACAGACATAGCTGACGGAGATCTTGACCCCTGAT 1392
Qy 515 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 534
Db 1393 CCTTCATTCCTGTTGCGGATGCGATTGCGGAGGCGAGCCCATTCCTCCGTCCTACCGA 1452
Qy 535 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeu 554
Db 1453 AGCGGCCACGCTGTGGTGTGCGGCATGTGTGACGATCCTGAAGGCGTTC-----TTC 1506
Qy 555 AspArgGlyGlyGluCysPheProAsnProValPheProSerAsp-----AspGly 571
Db 1507 GACTCCGCGCATCGAGATC-----GATCAGGTGTTCGAGGTGACAAAGATGAGGACAAG 1560
Qy 572 LeuGluLeuLeuAsnGlyAlaCysLeuThrTyrGluGlyGluIleAsnLysLeu 591
Db 1561 CTTGTGAAGTCGTCTTTCAAGGGAAT---CTCACTGTTCGCGGTGAATTGAACAAGCTC 1617
Qy 592 AlaValAsnValAlaPheGlyAtGlnMetLeuGlyIleHisTyrArgPheAspGlyIle 611
Db 1618 GCCGACAAATATTGCGATCGGCGGTAACATGCGAGGTGTCTACTACTTCTCTGACCACTC 1677
Qy 612 GlnGlyLeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThr 631
Db 1678 GAGTCACTTCTGCTCGTGAGCAGTTGCGATTGGAATCTTGGAGAGCAAGTCTGACG 1737
Qy 632 PheAlaGluAlaThrPheGluPheArgLeuPheThrGlyGluValIleLysLeu 650
Db 1738 TATGGCAGAACTTCTTCTTCAACTTCCGAAATTTGATGGAATACAAATCCAGATC 1794

RESULT 5
AAV56021
ID AAV56021 standard; cDNA; 1791 BP.
XX AC
XX AAV56021;
XX
DT 09-DEC-1998 (first entry)
XX
DE Haloperoxidase, CP.BP02 encoding cDNA.
XX
XX Haloperoxidase; enzyme; CP.BP01; CP.BP02; vanadium; halide specificity;
XX bromine; iodine; ds.
XX
OS Corallina pilulifera.
XX
XX Key Location/Qualifiers
XX CDS 1..1791
XX /*tag= a
XX /product= "Haloperoxidase, CP.BP02"
XX /note= "the stop codon is not indicated"
XX
XX JP10248581-A.
XX
XX 22-SEP-1998.
XX
XX 06-MAR-1997; 97JP-00070539.
XX
XX 06-MAR-1997; 97JP-00070539.
XX
XX (SAKA ) OTSUKA PHARM CO LTD.
XX
XX WPI; 1998-560733/48.
XX
DR P-PSDB; AAW80551.

```

```

XX New haloperoxidase gene - and corresponding vector, transformed host cell
PT and method of preparation.
XX
XX Claim 2; Page 15-18; 20pp; Japanese.
XX
XX This cDNA encodes a haloperoxidase enzyme. The haloperoxidase genes
CC encoding the enzymes CP.BP01 and CP.BP02 are isolated from Corallina
CC pilulifera. A host cell transformed with a vector containing the
CC haloperoxidase genes can be used for the recombinant production of the
CC enzyme. The haloperoxidases of the invention are vanadium-dependent and
CC have halide specificity of bromine and iodine
XX
SQ Sequence 1791 BP; 390 A; 459 C; 507 G; 435 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 7,53e-48 Length: 1791
Score: 614.50 Matches: 193
Percent Similarity: 43.04% Conservative: 70
Best Local Similarity: 31.59% Mismatches: 223
Query Match: 17.42% Indels: 125
DB: 2 Gaps: 21

US-10-691-383-2 (1-676) x AAV56021 (1-1791)
Qy 141 ArgAspAsnValAlaPheAlaSerArgAspAlaAlaArgArgGluArgAspGlyThr 160
Db 31 CGCGCCCAAGGCTTCATTCGATACGCGTGTAGTCGCGCGAGCTGGCACTC---GCCGT 87
Qy 161 GlyThrValCysGlnIleThrAsnGlyGlu----- 170
Db 88 GGAGTGTGTCATCGTTGTCAAATGGAGAGGAGCTTCTTACCCCAACTGTGAGACCGGC 147
Qy 171 ---ThrAspLeuAlaThrMetPheHisLysSerLeuProHisAspGluLeuGlyGln--- 188
Db 148 GATCCAAGCTTCATTCGTAGCTTCACAAAGGCTTCCACATGACGACATGGCGCTATT 207
Qy 189 ValThrAlaAspPheAlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIle 208
Db 208 ATTAGTCCCGACGATTTCTTGGCCTTCGTCGTCGATCAATAGTCGAGATGAGAAGGAG 267
Qy 209 CysGluAspVal-----ProAlaGlyAspProAlaGlyArgLeu----- 221
Db 268 ATCCCGCACCTTACATTTGGGGCCAGCTTCGACCCCTGAGACTGGCTTACCNAATCTGGCGC 327
Qy 222 -----ValAsnProThrAlaAla 227
Db 328 TCAGATCTGGCGAATTCCTCTCGAGCTCGAAGTGGCGAGATGGGAAACAGCTCTGCCGT 387
Qy 228 PheAlaIleAspIleSerGlyProAlaPheSerAlaThrThrIleProProValProThr 247
Db 388 CTTACCTTCGACCTGGAGGCGCCGACGCGCAGTCGGTGGCCATGCGCACCGCGCTGTG 447
Qy 248 LeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAlaArg 267
Db 448 CTCATGAGCCCTGAGCTCATCGCGGAGATGCGCAACTGTACTGTATGGCGCTTGGACGC 507
Qy 268 AspValProPheMetGlnTyrGlyThrAspGlu-----IleThrThrAlaAla 284
Db 508 GACATCGAGTTAGCGAGTTCGAGTCCCAAGAACCGCAGCTTTTATTAGTCTGCTATT 567
Qy 285 AlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAla----- 298
Db 568 GAACGGCTTAACGGGTGAGTGTGTTCC---AACACACCGGCAAGCTTGGTGTATCCGCT 624
Qy 299 -----ValSerIleGlySerAspGlyThrValAspPro 309
Db 625 GCGGAAATCGTCCGCGTCCGCGGTGAGTGCACCTCGGAAC----- 666
Qy 310 PheSerGlnLeuPheArgAlaThrPheValGlyValGluThrPheValSerGln 329
Db 667 -----TTGTTCCCGCGGTATTCTTCCCGGTTCCTGAGGTGCGGCCCGGTATCTCAGCCAG 717

```

```
Qy 330 LeuLeu-----ValAsnSerPheThrIle----- 337
Db   :::          :::          :::          :::
Qy 718 TACATCATGTTGTAGCAAGCAGATTGGCTCTGCGACGGTTCGCAACAACAACTTTCGTG 777
Db   TACATCATGTTGTAGCAAGCAGATTGGCTCTGCGACGGTTCGCAACAACAACTTTCGTG 777
Qy 338 -----AspAlaIleThrValGlu 343
Db   -----AspAlaIleThrValGlu 343
Qy 778 AGCCCCAATGTCGCGATGAATTTGATGGTGAATCCCTACGGAAGCATCACCATTAGC 837
Db   AGCCCCAATGTCGCGATGAATTTGATGGTGAATCCCTACGGAAGCATCACCATTAGC 837
Qy 344 ProLysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPheAspGluTrpLeu 363
Db   ProLysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPheAspGluTrpLeu 363
Qy 838 CAGCGTGTGGTATCGCCACCGCTGGCGGACGATTCATGACTGATTTGAAGGTATTCCTT 897
Db   CAGCGTGTGGTATCGCCACCGCTGGCGGACGATTCATGACTGATTTGAAGGTATTCCTT 897
Qy 364 AsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluLeuArgPhe 383
Db   AsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluLeuArgPhe 383
Qy 898 GACGTCCAAGACGGTGCAGACTCCGAGGATTTGAGTCGTATGAGCCGGGACGCGCTC 957
Db   GACGTCCAAGACGGTGCAGACTCCGAGGATTTGAGTCGTATGAGCCGGGACGCGCTC 957
Qy 384 IleArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsnThrGluAla 403
Db   IleArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsnThrGluAla 403
Qy 958 ATTGGAGCAGATCCGCGATCTTGGACGCTGGGTGCACCTTTGACGCACTGTACGAGGCTAC 1017
Db   ATTGGAGCAGATCCGCGATCTTGGACGCTGGGTGCACCTTTGACGCACTGTACGAGGCTAC 1017
Qy 404 TyrArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGly 423
Db   TyrArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGly 423
Qy 1018 CTCAAATGCGTCCCTGATCTTGTGGCAACGGGTGCGTTCGATCC-----AACCTT 1071
Db   CTCAAATGCGTCCCTGATCTTGTGGCAACGGGTGCGTTCGATCC-----AACCTT 1071
Qy 424 ProPheIleAspSerAspArgGlnAlaGly-----PheValAsnPheGlyThrSer 440
Db   ProPheIleAspSerAspArgGlnAlaGly-----PheValAsnPheGlyThrSer 440
Qy 1072 CCTTTCAGCAGGAGGACAACCTCGACACCATGATGTGTTGTGAACCTTCGGATCCGCA 1131
Db   CCTTTCAGCAGGAGGACAACCTCGACACCATGATGTGTTGTGAACCTTCGGATCCGCA 1131
Qy 441 HisTyrPheArgIleIle-----GlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGln 459
Db   HisTyrPheArgIleIle-----GlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGln 459
Qy 1132 CACGTGCTGATCTGTGCTGACTGAAGTGGCTACGCGCGCTTGAAGCGGTACGCTATCAG 1191
Db   CACGTGCTGATCTGTGCTGACTGAAGTGGCTACGCGCGCTTGAAGCGGTACGCTATCAG 1191
Qy 460 LysTrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThrLeu----- 476
Db   LysTrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThrLeu----- 476
Qy 1192 AAGTTTAACATTCATCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 1251
Db   AAGTTTAACATTCATCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 1251
Qy 477 ---HisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsn 495
Db   ---HisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsn 495
Qy 1252 AAGAAATCTTCTTGGCGGTAGT-----GATATAATATTCCTCGAAGTT 1296
Db   AAGAAATCTTCTTGGCGGTAGT-----GATATAATATTCCTCGAAGTT 1296
Qy 496 AspGluLeuLeuLysArg-----ValAlaGluIleAsnAlaAla 508
Db   AspGluLeuLeuLysArg-----ValAlaGluIleAsnAlaAla 508
Qy 1297 AGTGAATGTCGAGGAGCTCTCATCAATTTCTGATGACGTTGCTGAGACAATGAAAA 1356
Db   AGTGAATGTCGAGGAGCTCTCATCAATTTCTGATGACGTTGCTGAGACAATGAAAA 1356
Qy 509 GlnAsnProAsnGlnAla-----ThrTyrLeuLeuProGlnAlaIle 523
Db   GlnAsnProAsnGlnAla-----ThrTyrLeuLeuProGlnAlaIle 523
Qy 1357 CAGAACAGGGCTGACGGGATGTGAGCCCGGATTAATCATTTCTGTGCGGATGCGCATTT 1416
Db   CAGAACAGGGCTGACGGGATGTGAGCCCGGATTAATCATTTCTGTGCGGATGCGCATTT 1416
Qy 524 GlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAla 543
Db   GlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAla 543
Qy 1417 GCCGAGGCGACCCATTCCTCCGCTCTATGGAAGTGGCCACGCTGTTGCTGGCGCA 1476
Db   GCCGAGGCGACCCATTCCTCCGCTCTATGGAAGTGGCCACGCTGTTGCTGGCGCA 1476
Qy 544 PheAlaThrValLeuLysAlaLeuIleGly-----LeuAspArgGlyGlyGlu 559
Db   PheAlaThrValLeuLysAlaLeuIleGly-----LeuAspArgGlyGlyGlu 559
Qy 1477 TGTGTGCAATCTCTCAAGCGGTTCTTCAGCGCCAACTTCAGATCCGATAAGGTGTTGAG 1536
Db   TGTGTGCAATCTCTCAAGCGGTTCTTCAGCGCCAACTTCAGATCCGATAAGGTGTTGAG 1536
Qy 560 CysPheProAsnProValPheProSerAspAspGlyLeuGluLeuIleAsnPheGlyGly 579
Db   CysPheProAsnProValPheProSerAspAspGlyLeuGluLeuIleAsnPheGlyGly 579
Qy 1537 -----GTCGACACTGATGAGGACAAGCTTTGAGTCTCTTTCAAGGGA 1581
Db   -----GTCGACACTGATGAGGACAAGCTTTGAGTCTCTTTCAAGGGA 1581
Qy 580 AlaCysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArg 599
Db   AlaCysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArg 599
Qy 1582 ACT---CTCACTGTGCGGTGAATTCGACACGCTCGCCGACATGTTGCGATCGGCGG 1638
Db   ACT---CTCACTGTGCGGTGAATTCGACACGCTCGCCGACATGTTGCGATCGGCGG 1638
Qy 600 GlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThr 619
Db   GlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThr 619
Qy 1639 AACATGCGGGGTCTACTACTCTCTGACGAGTTCGAGTCTGCTACTGCTGCTGCTGCTG 1698
Db   AACATGCGGGGTCTACTACTCTCTGACGAGTTCGAGTCTGCTACTGCTGCTGCTGCTG 1698
Qy 620 IleThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluAlaThrPheGlu 639
Db   IleThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluAlaThrPheGlu 639
Qy 1699 ATTGCGATGGATCTTGGAGGACGAGAGCTGACGATGCGGAGAACTTCTTCTTCAAT 1758
Db   ATTGCGATGGATCTTGGAGGACGAGAGCTGACGATGCGGAGAACTTCTTCTTCAAT 1758
Qy 640 PheArgLeuPheThrGlyGluValIleLysLeu 650
```

Db 1759 TTGCCGAATTCACGGAACACTACAATCCAGATT 1791

## RESULT 6

AAD54222 standard; DNA; 9579 BP.

XX AAD54222;

XX 17-JUN-2003 (first entry)

XX Streptomyces platensis subspecies rosaceus dorrigocin ORF5 DNA.

XX Polyketide biosynthesis; dorrigocin; DORR; lactimidomycin; LACT; gene; ds.

XX Streptomyces platensis.

XX Key Location/Qualifiers

XX CDS 1..9579

XX FT /tag= a

XX FT /product= "ORF5 protein"

XX WO200288176-A2.

XX 07-NOV-2002.

XX 26-APR-2002; 2002WO-CA000591.

XX 26-APR-2001; 2001US-0286346P.

XX (SCOP-) ECOPIA BIOSCIENCES INC.

XX Farnet CM, Zazopoulos E, Staffa A, Yang X;

XX WPI; 2003-201222/19.

XX P-PSDB; AAE35488.

XX Novel isolated or purified polypeptide involved in biosynthesis of polyketide dorrigocin or polyketide lactimidomycin, useful for preparing dorrigocin or lactimidomycin.

XX Claim 6; Page 137-142; 312pp; English.

XX The invention relates to novel proteins involved in the biosynthesis of polyketide dorrigocin (DORR) or lactimidomycin (LACT) biosynthesis by microorganisms. Sequences of the invention allow direct manipulation of dorrigocin, lactimidomycin and related chemical structures via chemical engineering of the enzymes involved in the biosynthesis of dorrigocin and lactimidomycin. They are useful for introducing chemical handles into normally inert positions that permit subsequent chemical modifications and facilitate the development of polyketides. The genes and proteins of the invention can also be used to generate a focused library of analogues around a polyketide lead candidate to fine-tune the compound for optimal properties. They are useful for generating antibodies specific for the polyketide biosynthesis. The present sequence is Streptomyces platensis subspecies rosaceus DORR ORF5 DNA

XX Sequence 9579 BP; 1344 A; 3657 C; 3384 G; 1194 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.:	0.323	Length:	9579
Score:	131.00	Matches:	152
Percent Similarity:	30.12%	Conservative:	82
Best Local Similarity:	19.56%	Mismatches:	254
Query Match:	3.71%	Indels:	289
DB:	7	Gaps:	38

US-10-691-383-2 (1-676) x AAD54222 (1-9579)

Qy 5 AlaAlaAspThrThrArgGly---SerProMetPro-AspThrGlyValleuArgIleuLe 23

Db 6726 GCCGCGATGAGCAGCGCGCCCGCCGACCTTCGACCGGAGCAGACTTCTACCGCGCTCGG 6785



```
Db      8700 AGTCTGAAGGTGCTGTCCCATG---CCTCAGCGGACTGCCGCGCCTTGCACTG 8756
|||||
Qy      563 -----AsnProValPheProSerAspGlyLeuGluLeuLe 575
|||||
Db      8757 CGACCGGCTCAACCGGACCTGCGCTCGAGCGGCGGATTGCGAGTCGTA 8805
|||||

RESULT 7
AAD54217
ID      AAD54217 standard; DNA; 52101 BP.
AC      AAD54217;
DT      17-JUN-2003 (first entry)
DE      Streptomyces platensis subspecies rosaceus dorrigocin DNA.
XX      Polyketide biosynthesis; dorrigocin; DORR; lactimidomycin; LACT; gene;
KW      ds.
XX      Streptomyces platensis.
FH      Key
CDS     Location/Qualifiers
FT      complement(67..3720)
FT      /*tag= a
FT      /product= "ORF1 protein"
FT      4032..5691
FT      /*tag= b
FT      /product= "ORF2 protein"
FT      /notes= "No start codon"
FT      /partial
FT      5767..6018
FT      /*tag= c
FT      /product= "ORF3 protein"
FT      6023..7993
FT      /*tag= d
FT      /product= "ORF4 protein"
FT      8009..17587
FT      /*tag= e
FT      /product= "ORF5 protein"
FT      17634..41714
FT      /*tag= f
FT      /product= "ORF6 protein"
FT      /notes= "No start codon"
FT      /partial
FT      41772..47633
FT      /*tag= g
FT      /product= "ORF7 protein"
FT      /notes= "No start codon"
FT      /partial
FT      47635..49890
FT      /*tag= h
FT      /product= "ORF8 protein"
FT      49922..50938
FT      /*tag= i
FT      /product= "ORF9 protein"
FT      51234..52079
FT      /*tag= j
FT      /product= "ORF10 protein"

W0200288176-A2.
XX
XX
XX      07-NOV-2002.
XX
XX      26-APR-2002; 2002WO-CA000591.
XX
XX      26-APR-2001; 2001US-0286346P.
XX      (ECOP-) ECOPIA BIOSCIENCES INC.
XX
XX      Farnet CM, Zazopoulos E, Staffa A, Yang X;
XX      WPI; 2003-201222/19.
DR
```

```
DR      P-PSDB; AAE35484, AAE35485, AAE35486, AAE35487, AAE35488, AAE35489,
DR      AAE35490, AAE35491, AAE35492, AAE35493.
XX
PT      Novel isolated or purified polypeptide involved in biosynthesis of
PT      polyketide dorrigocin or polyketide lactimidomycin, useful for preparing
PT      dorrigocin or lactimidomycin.
XX
PS      Claim 1; Page 85-113; 312pp; English.
XX
CC      The invention relates to novel proteins involved in the biosynthesis of
CC      polyketide dorrigocin (DORR) or lactimidomycin (LACT) biosynthesis by
CC      microorganisms. Sequences of the invention allow direct manipulation of
CC      dorrigocin, lactimidomycin and related chemical structures via chemical
CC      engineering of the enzymes involved in the biosynthesis of dorrigocin and
CC      lactimidomycin. They are useful for introducing chemical handles into
CC      normally inert positions that permit subsequent chemical modifications
CC      and facilitate the development of polyketides. The genes and proteins of
CC      the invention can also be used to generate a focused library of analogues
CC      around a polyketide lead candidate to fine-tune the compound for optimal
CC      properties. They are useful for generating antibodies specific for the
CC      polyketide biosynthesis. The present sequence is Streptomyces platensis
CC      subspecies rosaceus DORR DNA
XX
SQ      Sequence 52101 BP; 7364 A; 20113 C; 17894 G; 6730 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      3.37      Length:      52101
Score:          131.00      Matches:      152
Percent Similarity: 30.12%      Conservative: 82
Best Local Similarity: 19.56%      Mismatches: 254
Query Match:     3.71%      Indels:      289
DB:             7      Gaps:       38

US-10-691-383-2 (1-676) x AAD54217 (1-52101)
Qy      5 AlaAlaAspThrThrArgGly---SerProMetPro-AspThrGlyValLeuArgLeuLe 23
|||||
Db      14734 GCCGAGATGAGCAGCGCGCGCCGACCTCTCGACGGGACGACTTCTACGCGGCTCGG 14793
|||||
Qy      23 uThrSerGluGlnArgAlaLysGlyTrpArgArgGlnLeuGluGlyGlyLysSerLeuG 43
|||||
Db      14794 CACCGCGGGCTCGACGCGCGCGGTACGCGCGCAGCGTCAAGGGGTCAACCGAATCGA 14853
|||||
Qy      43 yPheHis-----ProSer-----GluThrProTyrIleLe 53
|||||
Db      14854 CGCCACCGGCTGCTGTACGGGTGCGCCGACCGGCATGTGCGAGCACCGACACAGCA 14913
|||||
Qy      53 sTyrLeuGluGlySerGluThrTrpLysLysValLysLeuProThrAspGlyIleSer-- 72
|||||
Db      14914 GCACGTGCAT---CTCCGCGCTGGGCGCTGCTCGGCTG---ACCCAGGGTGTTCAGCA 14967
|||||
Qy      73 -----AlaSerLysIleLeuGlyLysIleMe 81
|||||
Db      14968 CGCGTGGGCGCGGCGCGACGCCCGCTGGTGGCGGTGCGATTCGTCAGCGCGGAGAGTG 15027
|||||
Qy      81 tAlaArgValArgIleAlaThrAlaLeuAlaValLeuAlaAlaProCysLeuAlaPh 101
|||||
Db      15028 GGAGCGCACCCGG-----GGATCTGCTGCGCGCG----- 15058
|||||
Qy      101 eAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGly----- 118
|||||
Db      15059 -----ACGTCCGACGCCGCTCTTC-----CATCGCGCTTTCCTCGA 15093
|||||
Qy      119 -----GluGlyArgHisLeuGlnThrCysThrAsnSerAsp-----AspAlaLeuAs 134
|||||
Db      15094 CGAGGACGCGCGCTGCTGGGCGGCTCGAGCGCGGCTTCACCGCGGCGCACCTGGA 15153
|||||
Qy      134 pProThrAlaProAsnArgArgAspAsnValAlaPheAla----- 147
|||||
Db      15154 GCCGGCACTCCCGGTGAGCGCGGACCGGCACTCGTGGCACTGCCGCGAGCGTCCGCTCC 15213
|||||
Qy      148 -----SerArgAspAlaAlaArgArgL 156
|||||
```

Db 15214 GGTCTGGAGACCGCGGTGGTGTACGCGGGAGTGCAGAGTCCGAGGCGCGTGCAGCGCGA 15273  
Qy 156 uArgAspGlyThrGlyThrValCysGlnIleThrAsnGly----- 169  
Db 15274 GCGCGAGCGGTCCGTGACCGTGGCGGGTTCGCGGACGCGCGCGCGTCTGTCGGGTC 15333  
Qy 170 -----GluThrAspLeuAlaTh 175  
Db 15334 GCTCGCGGAGACCGTCCGCGACCTGCTCAAGTTTCGACCTGGCGGACATCGACCTCGACAC 15393  
Qy 175 xMetPheHisLys---SerLeuProHisAspGluLeuGlnValThrAlaAspPhe 194  
Db 15394 GCACTTCACCGGTACCGTTCGAGTTCATCGCGCTGCCAATCACTCGGAACTCGAACTCAA 15453  
Qy 194 eAlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAspValProAl 214  
Db 15454 CGGCGTCTCCGCGACGACCTCACCCCGCGCGTCTTCTCGAGTGTCTCCGAC----- 15505  
Qy 214 aGlyAspProAlaGlyArgLeuValAsnProThrAlaAlaPheAlaIleAspIleSerGl 234  
Db 15506 -----ATCCGCGCTCCGCGAGTACCTGCTCGACCGCTACGG 15543  
Qy 234 yProAlaPheSerAlaThrThr-----IleProValProThrLeuSe 249  
Db 15544 CCGGAGCTGAGCTCCCGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15603  
Qy 249 rSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAlaArgAspVa 269  
Db 15604 GTCCCCA-----GT 15612  
Qy 269 lProPheMetGlnTyrGlyThrAspGluIleThrThrAlaAlaAlaAsnLeuAlaGl 289  
Db 15613 GCCGATCCGCGACCG 15663  
Qy 289 yMetGlyGlyPheProAsnLeuAspAlaValSerIleGlySerAspGlyThrValAspPr 309  
Db 15664 TCGCGAGCGGTTCCTCCCGCGCGGAC-----GACCTGGGAC 15699  
Qy 309 oPheSerGlnLeuPheArg-----AlaThrPheValGly-----Va 321  
Db 15700 CTTCGCGAGCAGCTGCGCGCGCGCGGAGGAGCTGATCGCGGACTACCCCGCGCGCGCTT 15759  
Qy 321 lGluThrGlyProPheValSerGlnLeuValAsnSerPheThrIleAspAlaIleTh 341  
Db 15760 CGACGGGGCGCCCTACCGCGAGGTCTGCGCGCGCGCGGAGCTTCCCGAAGTTTCCGCGCG 15819  
Qy 341 rValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPheAspGl 361  
Db 15820 GATCGAG---GGCGTGAGCCGCTTCGACGCGGACTTCTCCACCTGTCGCGGTGGAGGC 15876  
Qy 361 u-----TrpLeuAsnIleGlnAs 367  
Db 15877 GGAGCTGATGACCCCGCAGCACCGGTGGCCCTGGAGACCGTGTGGCGCGCGTGGAGAA 15936  
Qy 367 nGlyGly-----ProProAlaGlyProGluGluLeu-----As 378  
Db 15937 CGCGCGGTACCGCGCGCGCGCGCTCCCGGAGAACCGCGGCTCTACTTCGCGGTCTCCGG 15996  
Qy 378 pGluGluLeuArgPheIleArgAsnAlaArgAspLeuAla----- 391  
Db 15997 CAGCGACTACACACCTGCTCAACGCGGAGTGGCGTGGCACCCGAGCGGTTCACCGCCAC 16056  
Qy 392 -----ArgValSerPheValAspAsnIleAsn----- 400  
Db 16057 CGGCAACGCCCACTCGATGTGGCCAAACCGGATCTCTAGCTCTGAGCGTGCACGGGCC 16116  
Qy 401 -----ThrGl 402  
Db 16117 GAGGAAACCGTCGACACGCGCTGCTCCAGTCTGCTCGCTGTCGCGGTGACCGCGCGTCGA 16176  
Qy 402 uAlaTyrArg-----GlySerLeuIleLeuGlu 412  
Db 16177 GCACATCCGTCGCGCGCGGATGCGAGATGGCCATCGCGCGCGGTGTCAACCTGCTGTGAG 16236

Qy 412 uLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSerAspArgGlnAl 432  
Db 16237 CGTGGACACCTTCGCGCGGACACATGCGGGC---ATGCTCAGCCCGACGCGCGGTG 16293  
Qy 432 aGlyPheValAsnPheGlyThrSerHisTyrPheArgLeuIleGly---AlaAlaGluLe 451  
Db 16294 CAAGACCTTCGCGCGCGCGCGGACCGGTACGTCCTCCGAGGCGCTGCGCGCGGTCT 16353  
Qy 451 uAlaGlnArgAlaSerCysTyrGlnLys-----TrpGlnValHisArgPhe 466  
Db 16354 GCTCAAGCCCTGCCCGACGCGGAGCGGAGCGGACGCCATCTGGGGCGTCTGCCGGGG 16413  
Qy 466 eAlaArg----- 468  
Db 16414 CAGCGCGAGAACACCGCGCGCGCGCGGTTCGCTGACCGCCCAACGCAAGGCGCA 16473  
Qy 469 -----ProGluAlaLeuGlyGlyThr-- 475  
Db 16474 GGCGCGCTGATCCAGGACGCCATCGCGCGCATCGACCCGACACATCGGCTACGTCGA 16533  
Qy 476 -LeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAs 495  
Db 16534 GGCGACCGGACGCGGACCGCGC---CTGGGAGACCGGTGAGGTCAACGCCCTCGACAG 16590  
Qy 495 nAspGluLeuLeuLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluVa 515  
Db 16591 CGCTACTCGCGCTCGCGACCGCGAG----- 16618  
Qy 515 lThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSe 535  
Db 16619 -----GGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTG 16647  
Qy 535 r-----GlyHisAlaThrGlnAsnGlyAlaPheAlaTh 546  
Db 16648 CGCGCTCGGTGCTGAGAACCAACATCGCGCGCGCGGAGTGGCGCGCGCGCTGGCGCG 16707  
Qy 546 rValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCysPhePro----- 562  
Db 16708 AGTGCTGAGGTGCTGCTGCCATG---CGTCACCGCGAGTGGCGCGCGCGCTTGCACGT 16764  
Qy 563 -----AsnProValPheProSerAspAspGlyLeuGluLeuLe 575  
Db 16765 CGACCGGCTCAACCCGCGACCTGCGCTCGAGCGCGGATTCGAGGTCTGA 16813

RESULT 8  
ADA30090  
ID ADA30090 standard; DNA; 11679 BP.  
XX ADA30090;  
XX  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE DNA encoding Acinetobacter baumannii protein #1377.  
XX  
XX ds; gene; Acinetobacter baumannii; bacterial disease; antibacterial;  
XX vaccine; plant biocontrol agent.  
XX  
XX Acinetobacter baumannii.  
XX  
XX US65G2958-B1.  
XX  
XX 13-MAY-2003.  
XX  
XX 04-JUN-1999; 99US-00328352.  
XX  
XX 09-JUN-1998; 98US-0088701P.  
XX  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX  
XX Breton G, Bush D;  
XX  
XX WPI; 2003-576092/54.

DR P-PSDB; ADA34216.

XX New Acinetobacter baumannii proteins and nucleic acids, useful as reagents  
 PT for diagnosing a bacterial disease, as components of antibacterial  
 PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for  
 PT plants.

XX Example; SEQ ID NO 1377; 328pp; English.

XX The invention relates to isolated Acinetobacter baumannii nucleic acids.  
 CC The A. baumannii nucleic acids and polypeptides are useful as reagents  
 CC for diagnosing a bacterial disease, as components of antibacterial  
 CC vaccines, as targets for antibacterial drugs, to detect the presence of  
 CC A. baumannii and other Acinetobacter species in a sample, in screening  
 CC compounds for the ability to interfere with the A. baumannii life cycle  
 CC or to inhibit A. baumannii infection, and as biocontrol agents for  
 CC plants. The present sequence represents DNA encoding an A. baumannii  
 CC protein.

XX Sequence 11679 BP; 3369 A; 2576 C; 2877 G; 2857 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	Score:	Length:
0.588	11679	
129.50	158	
30.41%	71	
Best Local Similarity:	245	
20.98%	279	
Query Match:	36	
3.67%		
8		

US-10-691-383-2 (1-676) x ADA30090 (1-11679)

QY 46 ProSerGlnThrProTyrIleLysTyrLeuGluLysSerGlu----- 59

DB 988 CCTGGTTCACAGTCAGTGTGACTATCCGACGGTAGTAGCAAACTGTTGTAGCAGGA 1047

QY 60 -----ThrTpyLysValLysValLysLeuProThrAspGlyIleSerLysIle 76

DB 1048 CCGGATGACTTGG-----ACTGTGCCAAACCCAGGTGTGATGATGGCGATGA 1098

QY 77 LeuGlyLysIleMetAlaA-gValArgIleAlaThr----- 88

DB 1099 GTT-----ACAGCTGTAGCAACAGACAGACCCCTGCTGGCAATACATCAGGC 1140

QY 89 ---AlaLeuAlaValValLeuAla---AlaProCysLeuAlaPheAspGluVal----- 104

DB 1141 CCAGCAACGGCAGGTGTGATGAGTTCGCGCAGACAGTACCTTGATGATGACTCACT 1200

QY 105 -----ThrAlaSerGlyVal--- 109

DB 1201 AACGACAGCACACAGCAGTCTTACTGTGTACAGTGAACGATCCGACTGCACTGTAGTTGTC 1260

QY 110 -----PheProGluGluHisLysHisThrGly----- 118

DB 1261 AATGTGATGGTGTGATATCCGCGAGTGAACAATGTTGACGCGCAGCTGGACGCTTGCA 1320

QY 119 -----GluGlyArgHisLeuGlnThrCysThrAsnSerAsp 130

DB 1321 GACAATACATTCCAAACACTCGCAGACGGTCCACACACCACTTACCGTACTGCAACAGAT 1380

QY 131 AspAla-----LeuAspProThrAlaProAsn 139

DB 1381 GCAGCAGTAACGTAGGTACAGATACAGGTGTTGTGACAGTGTGATACAGCTGCTCCAAAT 1440

QY 140 -----ArgArgAspAsnValAlaPheAlaSer 148

DB 1441 ACAGCTGGCTTACCTTTTACGATTGACTCAGTAACAGCTGACAACTGTGATTAATGATCA 1500

QY 149 ArgArgAspAlaAlaArgArg-----GluArgAsp 158

DB 1501 -----GAAGCAGCAGGTAAATGTCACGATTACTGGTGTCTGAAAAACATTCGACGAGAT 1554

QY 159 GlyThrGlyThr---ValCysGlnIleThrAsnGlyGluThrAspLeuAlaThrMet--- 176

DB 1555 GCAGCTAAACACAGCGCGTTCACAGTTGTTATATGGGGTAACTTATATGCAACGGTAGAT 1614

QY 177 -----PheHisLysSerLeuProHisAspGluLeuGlyGlnValThrAla 191

DB 1615 AAAACAGCAGCGCAGTGGACAGTAAGTGTGCGGGTAGTGGTTG----- 1659

QY 192 AspAspPheAlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAsp 211

DB 1660 -----GTGCTGATGCAGATAAAGCATGTGCTGCTAAA 1692

QY 212 ValProAlaGlyAspProAlaGlyArg-----LeuValAsnProThrAlaAlaPheAla 229

DB 1693 GTAAAGTTTACAGATGCAGCAGGTAAATAGCAGCACTGTTAAACGATACGCAAAATTTATACA 1752

QY 230 IleAspIleSerGlyProAlaPheSerAlaThrIleProValProValProThrLeuSer 249

DB 1753 TTAGACACAGCTGCTCTCTGCA-----GCGCCAGTAATGACCCAGTT----- 1794

QY 250 SerProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAlaArgAspVal 269

DB 1794 ----- 1794

QY 270 ProPheMetGlnTyrGlyThrAspGluIleThrThrAlaAlaAlaAsnLeuAlaGly 289

DB 1795 -----AACGGGACAGACCCCAATTTACAGGTACAGCAGAACCTGGTTCAACAGTA 1842

QY 290 MetGlyGlyPheProAsnLeuAspAlaValSerIle-----GlySerAspGlyThrVal 307

DB 1843 ACAGTAACCTATCTTAATGTTGACAGCAACAGTGTGTAGCAGCCGCGGCGAGTTGG 1902

QY 308 Asp---ProPheSerGlnLeuPheArgAlaThrPheVal---GlyValGluThrGlyPro 325

DB 1903 TCAGTACCAACCCCTGGCTTAATGATGGTACGAGATTTGAGCAATGTCTACAGATCCA 1962

QY 326 PheValSerGlnLeuValAsnSerPheThrIleAspAlaIleThrValGluProLys 345

DB 1963 GCAGGCAACCCATCTTTGCCAGGTACAGCTACTGTTGATGCA-----GTTGGTCCAAAT 2016

QY 346 GlnGluThrPheAlaProAspLeuAsnTyrMetValAspPheAspGluTrpLeuAsnIle 365

DB 2017 ACCGAT-----GGTGTAACTTTACGGTGTGATTTCAGTAACAGCTGACAACTGTG 2064

QY 366 GlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluLeuArgPheIleArg 385

DB 2065 ATTAATGCATCAGACGCGTCAGCAAGCTTACTGTGTTACTGGTGTATTGAAAACGTT-- 2121

QY 386 AsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsnThrGluAlaTyrArg 405

DB 2122 ---CCGCGCAGATGCAGCAAAATACAGTGTGCTCACTGTTGTAATCAATGCCAGCGTATACT 2178

QY 406 GlySerIleLeuLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPhe 425

DB 2179 GCAACTGTA----- 2187

QY 426 IleAspSerArgGlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeu 445

DB 2188 -----GATAGCAGCAGCGC----- 2202

QY 446 IleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnValHisArg 465

DB 2203 -----ACATGCAGCAGTA----- 2214

QY 466 PheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeuAsp 485

DB 2215 -----AGCGTACCAGGTAGTACTTAAC 2238

QY 486 AlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLysArgValAlaGluIle 505

DB 2239 CGCGATGCAGAT-----AAGCAGGATTCGCTAAAGTAACGTTTACA 2280

QY 506 AsnAlaAlaGlnAsnPro-----AsnAsnGluValThrTyrLeuLeu----- 519

DB 2281 GATCGCGCAGGTAATAGCAGCAGTGTTAACGATACACAAACATATACAACTCGATACCAC 2340



QY 520 ---ProGlnala-----Ile 523  
DB 2341 GCACCTGATGCACCAAGTAATTAACCCGGTTAAGCGGACAGACCGGATTACAGGTACGGCA 2400  
QY 524 GlnValGlySerProThrHisProSerTyProSerGlyHisAlaThrGlnAsnGlyAla 543  
DB 2401 GAGCCTGGTTCAACAGTAAGTGTGACTTATCCAGATGCGAGTACAACA-----2448  
QY 544 PheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCysPheProAsn 563  
DB 2449 -----ACAGTTGTTGACGACCGGAT---GGCACTTGACAGTACCAAC 2490  
QY 564 ProValPheProSerAspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThr 583  
DB 2491 CCAGGTTTAAATGATGCGGAT-----2511  
QY 584 TyrGluGlyGluLeuAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGly 603  
DB 2512 -----AAAGTTACAGCAATGTGTACAGATCCAGCA---GGCAACCATCATACCA 2559  
QY 604 IleHisTyrArgPheAspGlyIleGlnGlyLeuLeuLeuGlyGluThrValArg 623  
DB 2560 GGTACAGCTACTGTGTGATGAGTGTGTCGCAATACCGATGGTGTATTTACAGGTTGAT 2619  
QY 624 ThrLeuHisGlnGluLeuMetThrPheAlaGluGluAlaThrPheGluPheArgLeuPhe 643  
DB 2620 TCAGTAACAGCTGACAAATGTGATTAATGATCATCAGAACATCAGGCAACGTTACTGTT---2676  
QY 644 ThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIleAspGlyAspMetCys 663  
DB 2677 ACAGGTGATTTGAAAAAGTTCCGGCAGATGCGCAATACAGTGTGCTACTGTTGTGATC 2736  
QY 664 SerGlyLeuValTyrThrGlyValAlaAspCysGlnAla 676  
DB 2737 AATGCCAGACGTAATCTGCAACTGTAGATAGACAGCA 2775

RESULT 9  
ACA43739

ID ACA43739 standard; DNA; 18930 BP.

AC ACA43739;

DT 19-JUN-2003 (first entry)

DE Prokaryotic essential gene #25396.

KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
drug design; gene.

OS Pseudomonas putida.

PN W0200277183-A2.

PD 03-OCT-2002.

PF 21-MAR-2002; 2002WO-US009107.

PR 21-MAR-2001; 2001US-00815242.

PR 08-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Foreyth RA, Xu HH;XX WPI; 2003-029926/02.  
DR P-PSDB; ABU99869.

XX New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 14; SEQ ID NO 31609; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX

SQ Sequence 18930 BP; 3570 A; 6691 C; 5875 G; 2794 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.:	3.79	Length:	18930
Score:	124.00	Matches:	147
Percent Similarity:	30.99%	Conservative:	56
Best Local Similarity:	22.44%	Mismatches:	226
Query Match:	3.51%	Indels:	229
DB:	7	Gaps:	31

US-10-691-383-2 (1-676) x ACA43739 (1-18930)

QY	111	ProGluGluHisLysHisThrGlyGluGlyArgHisLeuGln-ThrCysThrAsn----	128
DB	573	CCTGATCGGCAGCGGCACAGTGGTGGCGACGCACTTCAACGTGACGTTGAATGCGCC	632
QY	129	-----SerAspAspAlaLeuAspProThrAlaProAsnArgArgAspAsnValAlaPh	146
DB	633	GCAGATCAACACGCGCAAAACCTTGATGTGCTGACTGACGCGCGCGGTATGTCTCG--	690
QY	146	eAlaSerArgArgAspAlaAlaArgArgGluArgAspGlyThrGlyThrValCysGlnI1	166
DB	691	-----GCGCCTGGCGCTGTCTACTGCGCGGACGCACTGCACCATCTG-----	732
QY	166	eThrAsnGlyGluThrAspLeuAla-----ThrMetPheHisLysSe 180	
DB	733	-----GCGCGCAGCGACCTGCTATCAATGAGAGCGGTAACACCCCTTACCGTCTGTC	785
QY	180	rLeuProHisAspGlu-----LeuGlyGlnValTh 190	
DB	786	CGAGCCCGGCTCGACGCTTTCCGTGCGAGCGCAGGTGGCGTATTGTGCGCACTGGCGT	845
QY	190	rAlaAspAspPheAlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIleCysG1	210
DB	846	GGCGCGT-----GCGGACGCGCGAGTTCAGCATTTACCT 878	

QY 210 uAspValProAlaGlyAspProAlaGlyArgLeuValAsnProThrAlaAlaPheAlaI 230  
 DB 879 GCACCCCAACAGCGAC-----GCCAGCCCTGGAGATCAGCCGCCGACAGCGCGC 932  
 QY 230 eAspIleSerGlyProAlaPheSerAlaThrIleProProVal-----ProThrLe 248  
 DB 933 TGGCAATATCTCGCCAGTTCGCAGCATCACGGG---CCGGAGCTGCACAACCCAGATAC 989  
 QY 248 uSerSerProGluLeuAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAlaArgAs 268  
 DB 990 CACAGCCCTTGAACAACCGACCGACCTG-----CCCTTGGCCGATGG 1031  
 QY 268 pValProPheMetGlnTyrGlyThrAspGluIleThr-----ThrAlaAlaI 285  
 DB 1032 CGTCACCTTTCACCGGACCGGTGAGCTGTGTGTACCGTGCAGGTGCGCATGTCGCCG 1091  
 QY 285 aAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAlaValSerIleGlySerAspG1 305  
 DB 1092 TAACCTGATTGGCACCGGT-----GTGTCGTGTCGCGATGG 1127  
 QY 305 ThrValAspProPheSerGlnLeuPheArgAlaThrPheValGlyValGluThrGlyPr 325  
 DB 1128 C-----CTGTTACGCTCACCTCTCGCCTCGCCAGACC----- 1161  
 QY 325 oPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIleThrValGluProLy 345  
 DB 1162 -----AATGTGAAGCACTCGACGTGCTGCTGTGTGATGCCGC 1199  
 QY 345 sGlnGluThrPheAlaPro-----AspLeuAsnTyrMetValAspPheAspGluTr 362  
 DB 1200 CGGCAACAGCTCTGCCCATTCGATTCGATGCGCGGACATCACACCGCTGACGCAGT 1259  
 QY 362 pLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluGluLeuAr 382  
 DB 1260 AACCAATATTACAGTTGGTGTGTGACGG---CCTGGCCCTGAGCGGTGCGCGGACCGAGT 1318  
 QY 382 gPheIleArgAsnAlaArg-----AspLeuAlaArgValSerPheValAs 397  
 DB 1319 CCAGTGTGGAAGTCCGCGATGCCAATGCGACTGTGATCG---GCACGGGTGTGCGGTGCC 1377  
 QY 397 pAsnIleAsnThrGluAlaTyrArgGlySerLeuIleLeuLeuLeuGlyAlaPheSe 417  
 DB 1378 -AAC-----GGCACCTTCTGTATGACCTGACCTGCCCGCCAGCGC 1412  
 QY 417 rArgProGlyIleAsnGlyProPheIleAspSerAspArgGlnAlaGlyPheValAsnPh 437  
 DB 1413 ACAGCCGGCGGACAGTTGAGCTGTGTCAGACCGAC---CCAAGCGGTAACTGCTTCAGT 1469  
 QY 437 eGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGlu----- 450  
 DB 1470 GGCTACCGAGTACGATGATACACTGACCATGACCGCGCCGACAGCCCGAGCAATCTGGCCAT 1529  
 QY 451 -----LeuAlaGlnArgAlaSerCysTyrGlnIlystrpGlnValHi 464  
 DB 1530 CGATGCTGATGGCACCACTCTCACCGGTACCGCGCGCTGGCAGCGGTGGAGGTGCA 1589  
 QY 464 sArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGlyAspLe 484  
 DB 1590 T-----GACGCAACGGCAGCTGATTTGGCAGTCCCATTTGCCAACGC 1631  
 QY 484 uAspAlaAspPheAspIleSerLeu-----LeuGluAsnAspGluLeuLeuLyAr 501  
 DB 1632 CGATGGCAGCTTACGATCGAATCCCGCAAGCCACCGGCACTGCTGGACGT 1691  
 QY 501 gValAla----- 503  
 DB 1692 GGTGGCATCATGACGCGCGCGTTCTCTGCTGCCGACAGATCACCGCGCCGACAT 1751  
 QY 504 -----GluIleAsnAl 507  
 DB 1752 CACGCCCTCTGCCACCGACCGAAGTGGCGGTGAGCGCGGACGCGTATCACCGG 1811  
 QY 507 aAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySe 527

DB 1812 CCCTGCCGACCGCGGACGACCGCTGCGCATCGTGGCTGCCGACGGAACCGAATGGCGCAC 1871  
 QY 527 r-----ProThr-----HisProSerTyrProSerG1 536  
 DB 1872 TGGCGTGTGCGGCCGACCGCGCTGTTCAGCTTCAACCTCAACCCGCGCACAAAGTTGACGG 1931  
 QY 536 yHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspAr 556  
 DB 1932 CGAAGTGTGTCAA-----GCCACGGCGACC-----GATGC 1961  
 QY 556 gGlyGlyGluCysPheProAsnProValPhe-----ProSerAspAspGlyLeuGlu-- 573  
 DB 1962 CGCTGGCAACACATCGCCAGCAGCGAGTACCGCACCGGATATCGACGGGTTCGATAC 2021  
 QY 574 -----LeuIleAsnPheGluGlyAlaCysLeuThrTy 584  
 DB 2022 CACGCCACCGCGACCGCGACTGACTGTGTGATCGGCTCGCGGCGACCAACTCAGTGG 2081  
 QY 584 xGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyI1 604  
 DB 2082 TCAGGCGGAA----- 2091  
 QY 604 eHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThrIleThrValArgTh 624  
 DB 2092 -----GCCGGCTCTACCGTGCAGGTACGTGA 2117  
 QY 624 rLeuHisGlnGluLeuMetThrPheAlaGluGluAlaThrPheGluPheArgLeuPheTh 644  
 DB 2118 CGCGGACGCAACATC-----CTCGGCAC 2141  
 QY 644 xGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIle 658  
 DB 2142 CGGACCGTGCCT-----GCCGATGCGACCTTCACGTGC 2175  
 RESULT 10  
 AAZ55887  
 ID AAZ55887 standard; DNA; 68750 BP.  
 XX AAZ55887;  
 XX 15-SEP-2003 (revised)  
 DT 10-APR-2000 (first entry)  
 XX Sorangium cellulosum 68.75 kb contig.  
 DE Epithelone biosynthesis; type I polyketide synthase; taxol substitute;  
 KW anticancer; ds.  
 XX Polyangium cellulosum.  
 OS  
 FH Key Location/Qualifiers  
 CDS 1..1826  
 FT /\*tag= a  
 FT /partial  
 FT /product= "Partial Orf 1 protein (AAV58580)"  
 FT /note= "No initiation codon given in the specification"  
 FT complement(1900..3171)  
 FT /\*tag= b  
 FT /product= "Orf 2 protein (AAV58581)"  
 FT 3415..5556  
 FT /\*tag= c  
 FT /product= "Orf 3 protein (AAV58582)"  
 FT complement(5612..5992)  
 FT /\*tag= d  
 FT /product= "Orf 4 protein (AAV58583)"  
 FT 6226..6675  
 FT /\*tag= e  
 FT /product= "Orf 5 protein (AAV58584)"  
 FT 7610..11875  
 FT /\*tag= f  
 FT /product= "Type I polyketide synthase, EFOS A (AAV58573)"  
 FT 11872..116104  
 CDS







PF 21-MAY-2002; 2002CA-02387401.  
XX  
PR 21-MAY-2001; 2001US-0291959P.  
PR 03-DEC-2001; 2001US-0334604P.  
XX  
XX (ECOP-) ECOPIA BIOSCIENCES INC.  
XX  
XX Staffa A, Farnet CM, Zazopoulos E;  
PI  
XX WPI; 2002-692206/75.  
DR P-PSDB; ABG97437.  
XX  
XX New enediyne polyketide synthase polypeptides involved in synthesis of  
PT the warhead structure in enediyne compounds useful in design of new  
PT enediyne compounds and identification of new enediyne biosynthetic loci.  
XX  
XX Claim 1; Page 183-186; 310pp; English.  
PS  
XX The invention relates to purified/isolated polypeptides involved in  
CC synthesis of the warhead structure in enediyne compounds, or their  
CC fragments or sequences 75% homologous to them. The polypeptides comprise  
CC five families of proteins: PKSE (polyketide synthase), TEBC (thioesterase  
CC unique to enediyne biosynthetic loci), UNBL, UNBV and UNBU (unique to  
CC enediyne biosynthetic loci L, B and V). The proteins are isolated from 10  
CC different micro-organisms. Also included are the nucleic acids encoding  
CC the proteins, antibodies specifically binding the polypeptides, and a  
CC computer system/computer readable medium comprising the sequences  
CC referred to above. The polypeptides and polynucleotides are useful in  
CC genetic engineering to design new enediyne compounds (which are potential  
CC anti-cancer drugs) and identify new enediyne biosynthetic loci. They can  
CC also be used to identify enediyne biosynthesis genes/gene fragments from  
CC samples containing genomic DNA e.g. in biomass from an environmental  
CC source (such as a mixed microbial culture/population of organisms), pure  
CC cultures or genomic libraries obtained from a mixed population of  
CC organisms or from pure culture. The present sequence encodes an enzyme of  
CC the enediyne warhead structure biosynthetic pathway  
XX  
SQ Sequence 5820 BP; 846 A; 2105 C; 2071 G; 798 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 0.92 Length: 5820  
Score: 123.00 Matches: 146  
Percent Similarity: 31.73% Conservative: 59  
Best Local Similarity: 22.60% Mismatches: 249  
Query Match: 3.49% Indels: 192  
DB: 6 Gaps: 32

US-10-691-383-2 (1-676) x ABS78676 (1-5820)

Qy 4 HisAlaAlaPheThrThrArgGlySerProMetProAspThrGlyValLeuArgLeuLeu 23  
Db 1816 CAGCCGACGACATCTACCGCGCGCGCAACTG---TCCACCGCGCGCGACAGGTGCC 1872

Qy 24 ThrSerGluGlnArgAlaLys-----GlyTrpArgGlnLeuGluGlyGluLys 40  
Db 1873 ACCGACGTCCCGCAGCGCGCATCGTACCAGTTCGTCGCGCGGCTCGCGGTCTGAAG 1932

Qy 41 SerLeuGlyPheHisProSerGluThrProTyrIleLysTyrLeuGluGlySerGluThr 60  
Db 1933 AGCCTCGGCATCGAGCGCGCGCGGTCC-----ACGGGCGCACAGCCTTCGCGAGCTCAC 1986

Qy 61 -----TrpLysValLysLeuProThrAspGlyIleSerAlaSerLysIleLeu 77  
Db 1987 GCCCTGCACGTGGCGCGCGCTCACCGACGCGAGTGTCTCAACTGGCCAGATCCGC 2046

Qy 78 GlyLysIleMetAlaArgVal----- 84  
Db 2047 GGCAAGGTGATGGCGACCGCAGCGCGGCGGCCCATGGCGGATCGCGGCCACG 2106

Qy 85 ----ArgIleAlaThrAlaLeuAla----- 91  
Db 2107 CCGAGTGTCCCGCGGCGCTGGCGCGGCGGCGAGAGGTCTGTCATCGCGGGCTACACGCC 2166

Qy 92 -----ValValLeuAlaAlaProCysLeuAlaPheAspGluValThrAla----- 106  
Db 2167 CCGGACGACGACTGTCTTCCGACCGCGCGGAGCGATCAGCCGCGTGGTTCGCGCGTCC 2226

Qy 107 -----SerGly 108  
Db 2227 CGCGCGGAGGGGTACCGCGCGCGCATCAAGTCTTCGCACGCTTCACCTCCGCGCGC 2286

Qy 109 ValPheProGluGluHisLysHisThrGlyGluGlyArgHisLeuGln----- 124  
Db 2287 GTCTGCCCGCGCGCGGAGCGATGACCGGGGAACTCGCCGCGATCGACTTCGCGCGGCTC 2346

Qy 125 -----ThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArg 140  
Db 2347 GACCGCGCGCTCTCTCCACGCGTACCGGTGAGTCTCTGCACCGCGCGGAGACCTCGCG 2406

Qy 141 -----ArgAspAsnValAlaPheAlaSerArg---ArgAspAlaAlaArgArgGlu 156  
Db 2407 GATCTGTCTGCGCGACCGAGGTGCTTCTCGCGGTCTCGTTCGTCGCGCGCGCGGAGGTC 2466

Qy 157 ArgAspGlyThrGlyThrValCysGlnIleThrAsnGlyGlu----- 170  
Db 2467 GCCGAGCGCAGCACCTGTGTGATGAGGTGCGCGCGCGCGCGGTCTCACCGCGCTGCTC 2526

Qy 171 -----ThrAspLeuAlaThrMetPheHisLysSerLeuProHisAsp 184  
Db 2527 GGCACCATCGCGCGCGGTACCCCGGTGCTTTCGATCGACACCGCACGCTGACGCTCGCG 2586

Qy 185 GluLeuGlyGlnValThrAlaAspAspPheAlaIle-----LeuGluAspCysIle 201  
Db 2587 CCGGTGTGAAGTGTGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 2646

Qy 202 Leu-----AsnGlyAspPheSerIle----- 208  
Db 2647 CTGTTGACGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 2706

Qy 209 -----CysGluAspValPro-----AlaGlyAspProAlaGlyArgLeuValAsn 223  
Db 2707 AGCCCGTGGAGCGCGCGGTCTCTCCATCGCGCGGTGCTGTCGCGCGCGCGCGGTGCG 2766

Qy 224 ProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThrThrIlePro 243  
Db 2767 CCGCGCGAGCGCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 2817

Qy 244 ProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTipMet 263  
Db 2818 ---CTGACCTGCTGCGC-----AAGCTCGGTGCGGCGCGGTGCGGTGCGGTGCGGTG 2868

Qy 264 AlaLeuAlaArgAspVal---ProPheMetGlnTyrGlyThrAspGluIleThr----- 280  
Db 2869 GCGGTACCGCGCGCGCGGTCTCTCCATCGCGCGGTGCTGTCGTCGTCGTCGTCGTCG 2928

Qy 281 -----ThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsn 295  
Db 2929 CAGCTGTCTCAACGCGTACCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 2988

Qy 296 LeuAspAlaValSerIleGlySer---AspGlyThrValAspProPheSerGlnLeuPhe 314  
Db 2989 TTCGCGACCGTGTCTCTCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 3048

Qy 315 Arg-----AlaThrPheValGlyValGluThrGlyProPheValSer 328  
Db 3049 AAACCCCGCGCACACACCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 3096

Qy 329 GlnLeuLeuValAsnSerPheThrIleAspAlaIleThrValGluProLysGlnGluThr 348  
Db 3097 -----GTCCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 3138

Qy 349 PheAlaProAspLeuAsnTyrMetValAspPheAspGluThrLeuAsnIleGlnAsnGly 368  
Db 3139 ---TCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 3195

Qy 369 GlyProAlaGlyProGluLeuAspGluLeuArgPheIleArgAsnAlaArg 388

```
Db 3196 GGTCAACCA-----CTGCGGAGCGCTGCGC-----GCG 3225
Qy 389 AspleuAlaArgValSerPheValAspAsnIleAsnThrGluAlaIleArgGlySerLeu 408
Db 3226 GCATTGGCCAGCGCGGGCTGGGACGGT----- 3255
Qy 409 IleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSer 428
Db 3256 GTCCTCTCTCCCTGAACCGCCAGACAGCGCTTCCGCGACGTCGCGCTGCTCTCGACGCG 3315
Qy 429 AspArg-----GlnAlaGlyPhe 434
Db 3316 GCGCGCGCGGTGTCGGCGCCCAACGCGACGCGTTCGTCGTGGTGCACGCGTCTC 3375
Qy 435 ValAsnPheGlyThrSerHisTyPheArgLeu----- 445
Db 3376 GCGCGCTCGGGCTGGGAGACGCTCGGTGGAGGACCGCTGGCGCCGACACGATC 3435
Qy 446 IleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyGlnLysTrpGlnValHisArg 465
Db 3436 GTCGACCTCGCGGACCTCGGCGCG----- 3459
Qy 466 PheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeuAsp 485
Db 3460 ---GTCGACCCGAGCCCTCGACCGCGCGTG---TCCACTGTGGTACCGAAGTGGCG 3513
Qy 486 AlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuLysArgValAlaGluIle 505
Db 3514 GCGACGACCGACTTCAGGAGAGTCGCTACGACACCGCGGAGTCCGACGCGTCCCGAG 3573
Qy 506 AsnAlaAlaGlnAsnPro 511
Db 3574 CTCGCGCGCTGACCCCG 3591
```

## RESULT 13

ACF67367\_38

Continuation (39 of 57) of ACF67367 from base 3800001 (Phototaxibodus luminescens nucleoti  
WP Sequence split into 57 fragments LOCUS ACF67367 Accession ACF67367

Fragment Name	Begin	End
WP ACF67367_00	1	110000
WP ACF67367_01	100001	210000
WP ACF67367_02	200001	310000
WP ACF67367_03	300001	410000
WP ACF67367_04	400001	510000
WP ACF67367_05	500001	610000
WP ACF67367_06	600001	710000
WP ACF67367_07	700001	810000
WP ACF67367_08	800001	910000
WP ACF67367_09	900001	1010000
WP ACF67367_10	1000001	1110000
WP ACF67367_11	1100001	1210000
WP ACF67367_12	1200001	1310000
WP ACF67367_13	1300001	1410000
WP ACF67367_14	1400001	1510000
WP ACF67367_15	1500001	1610000
WP ACF67367_16	1600001	1710000
WP ACF67367_17	1700001	1810000
WP ACF67367_18	1800001	1910000
WP ACF67367_19	1900001	2010000
WP ACF67367_20	2000001	2110000
WP ACF67367_21	2100001	2210000
WP ACF67367_22	2200001	2310000
WP ACF67367_23	2300001	2410000
WP ACF67367_24	2400001	2510000
WP ACF67367_25	2500001	2610000
WP ACF67367_26	2600001	2710000
WP ACF67367_27	2700001	2810000
WP ACF67367_28	2800001	2910000
WP ACF67367_29	2900001	3010000
WP ACF67367_30	3000001	3110000
WP ACF67367_31	3100001	3210000
WP ACF67367_32	3200001	3310000

```
Qy 8 ThrThrArgGlySerProMetProAspThrGlyValLeuArgLeuLeuThrSerGluGln 27
Db 93960 TCTGCGCGAGCCAGCCGCTTCCAGAAATTGGG-----TGGAAAAATTCGNAAGG 94010
Qy 28 ArgAlaLysGlyTrpArgArgGlnLeuGluGlyLysSerLeuGlyLysPheHisProSer 47
Db 94011 TATTCTGATGTTGG-----GCCTCGGGGTAC-----GGG 94040
Qy 48 -GluThrProTyrIleLysTyrLeuGluGlySerGluThrTrpLysLysValLysLeuPr 67
Db 94041 TAAGACTTTACTGGCAAAAGCAATTGCTGGT-----GAAGCGAAAGTGCC 94085
Qy 67 oThrAspGlyIleSerAlaSer-----LysIleLeuGlyLysIleMetAlaAr 83
Db 94086 GTTCTTTACTATCTCTGGTTCTGATTTTGTGAATGTTTGTGGTGTGGTGCATCCCG 94145
Qy 83 qValArgIleAlaThrAlaLeuAlaValLeuAlaAlaProCysLeuAlaPhe---As 102
Db 94146 TGTTCGTGATATGTTTCGACAGCGG---AAGAAAGACGACCTTGATATCTTTATTGA 94202
Qy 102 pGluValThrAla-----SerGlyValPheProGluGluHi 114
Db 94203 TGAATTGATGCGGTGGCGCTCAGCTGGCGCTGTTTGGGTGGCGGTACGATGAACG 94262
Qy 114 sLysHisThr-----GlyGluGlyAr 121
Db 94263 TGAGCAGACTCTAAACCAGATGTTGTTGAGATGGATGGTTTGAAGCAACGAAGGTAT 94322
Qy 121 gHisLeuGlnThrCysThrAsnSerAspAlaLeuAspProThr-----AlaProAs 139
Db 94323 TATCGTTATTGCGCGGACTAACCGTCCGACGTAATTGATCCTCGGTGTTGCGTCCAGG 94382
Qy 139 nArgArgAspAsnValAlaPheAlaSerA-gArgAspAlaAlaA-gArgGluArg----- 157
Db 94383 TCGTTTGGACCGTCAGGTGTTGTTGTTGTTGCCAGACGTCAGTGGCGCGGAGCATTC 94442
Qy 158 -----AspGlyThrGlyThrValCysGlnIleThrAs 168
```

Db 94443 GAAAGTTTCATATGCGCGGGTTCCTCGGATACAGATGTTGTATGCTCTTCGTCATGCTCG 94502  
Qy 168 nGlyGlu-----ThrAspLeuAlaThrMetPheHisLysSerLeuProHi 183  
Db 94503 CGGTACTCCGGATTCCTGCTGCTGATTTGGCTAACTTGGTGAATGAAGCTGCTTTGTT 94562  
Qy 183 sAspGluLeuGln-----ValThrAlaAspPheAlaLeuLeuGluAspCy 200  
Db 94563 TGCTGCCCGTGTACAGACGCTAGTTCTATGGTTGAATTCGAAAGCGAAGACAA 94622  
Qy 200 sileLeuAsnGly-----AspPheSerIleCysGluAspVal----- 212  
Db 94623 AATCATGATGGCGCTCAACGCGCTCTATGGTATGATGACGGAAGAGCAAGATCAAC 94682  
Qy 213 -----ProAlaGlyAspPro-----AlaGlyArgLeuVal-----AsnProTh 225  
Db 94683 TGCTTATCATGAAGCAGCGCATGCTATATTATGGCGCTCTGTTCTCGAGCATGCTCTG 94742  
Qy 225 rAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThrThrIleProProVa 245  
Db 94743 TCATAAAGTTACCATTTATTCGGCGTGTGCGCTTTGGGGGTGACTTTTTCTTACCGGA 94802  
Qy 245 lPro-----ThrLeuSerSerProGluLeuAlaAlaClnLeuAlaGluLeuYrTr 262  
Db 94803 AGCGATCAGATCAGCGCTAGTCGCCAGAAAGTTGGAAGCCAGATTTCAACCTTATATGG 94862  
Qy 262 pMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGluIleThrThrTh 282  
Db 94863 AGTCTGCTGCGGGAAGAGATT-----ATTATGGCCACAGACAGTGTCTTCTACTGG 94913  
Qy 282 rAla-----AlaAlaAsnLeuAlaGlyMetGlyG1 292  
Db 94914 TGCATCTAATGATATCAAGTGGCGCATCAATTTGCCCTTAACATGTTAAGCAGTGGGG 94973  
Qy 292 yPhe-----ProAsnLeuAspAlaValSerIleGlySerAspGlyThrVa 307  
Db 94974 TTCTCGGAAAAAAGTTGGCCCATTTACTCTATGCA-----GAGGAAGAGGGGAAGT 95024  
Qy 307 lAspProPheSerGlnLeuPheArgAlaThrPheValGlyValGluThrGlyProPheVa 327  
Db 95025 GTTCTTGGCGGTTTCAGTTGCTAAAGCTAAGCAGATGCTGATGATAACTGCTGCTCTTAT 95084  
Qy 327 lSer-----GlnLeuLeuVa 332  
Db 95085 TGATCAGAAATTAAGCAATTATCGATCATTAATCAAGTCTCGTCAGATTCGAT 95144  
Qy 332 lAsnSerPheThrIle-----AspAlaIleThrValGluProLysGlnG1 347  
Db 95145 GGACAACCTTGATGTTCTGCAATTCGATTCGATAAGATGCAATTCATG-----AAGTACGA 95195  
Qy 347 uThrPhe-----AlaProAspLeuAsnTyrMetValAspPheAspGluTrpLeuAsnIleG1 366  
Db 95196 AACTATTGATGCTCCCGAGATT-----GATGACTTAATGAATCGTAC 95237  
Qy 366 nAsnGlyGlyProProAlaGlyProGluGlu-----LeuAspGluGluLeuArgPh 383  
Db 95238 ANATGTTGCTCCACCAGCAGATGGGAAGTATGATACGGCAGCAGTAAATACAGTCGCC 95297  
Qy 383 eileArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsnThrGluAl 403  
Db 95298 AACAGGTAATACACTCTCCACAAGCTGCAAAACCACTGATGATATCAATACATTGGC 95357  
Qy 403 aTyrArgGlySerLeuIle 409  
Db 95358 GCAAGTGGTGAGCAAGTA 95376

## RESULT 14

ACF65388\_09/c  
Continuation (10 of 13) of ACF65388 from base 900001 (Photorhabdus luminescens nucleotid  
WP Sequence split into 13 fragments LOCUS ACF65388 Accession ACF65388  
WP Fragment Name Begin End  
WP ACF65388\_00 1 110000  
WP ACF65388\_01 100001 210000

WP ACF65388\_02 200001 310000  
WP ACF65388\_03 300001 410000  
WP ACF65388\_04 400001 510000  
WP ACF65388\_05 500001 610000  
WP ACF65388\_06 600001 710000  
WP ACF65388\_07 700001 810000  
WP ACF65388\_08 800001 910000  
WP ACF65388\_09 900001 1010000  
WP ACF65388\_10 1000001 1110000  
WP ACF65388\_11 1100001 1210000  
WP ACF65388\_12 1200001 1225559  
Alignment Scores: 54 Length: 110000  
Pred. No.: 123.00 Matches: 121  
Score: 34.91% Conservative: 56  
Percent Similarity: 23.87% Mismatches: 191  
Best Local Similarity: 3.49% Indels: 139  
Query Match: 7 Gaps: 28  
DB: 7  
US-10-691-383-2 (1-676) x ACF65388\_09 (1-110000)  
Qy 8 ThrThrArgGlySerProMetProAspThrGlyValLeuArgLeuLeuThrSerGluGln 27  
Db 30976 TCTGCGCAGCCAGCGCTTCAGAAAGTTGGG-----TGGAAAAATTCGAAAGG 30926  
Qy 28 ArgAlaIleGlyTrpArgArgGlnLeuGluGlyLysSerLeuGlyPheHisProSer 47  
Db 30925 TATTCTGATGTTGG-----GCCCTCCGGGTAC-----GGG 30896  
Qy 48 -GluThrProTyrIleLysTyrLeuGluGlySerGluThrTrpLysLysValLysLeuPr 67  
Db 30895 TAAGACTTACTGCGCAACCAATTCGTGT-----GAACGGAAGTGCC 30851  
Qy 67 oThrAspGlyIleSerAlaSer-----LysIleLeuGlyLysIleMetAlaAr 83  
Db 30850 GTTCTTTACTATCTCTGTTCTGATTTTGTGAAATGTTTGTGTTGTTGTTGTCATCCCG 30791  
Qy 83 gValaArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPhe 102  
Db 30790 TGTTCGTGATATTCAGCAGCG-----AAGAAAGCAGCACCTTGTATCATCTATTGA 30734  
Qy 102 pGluValThrAla-----SerGlyValPheProGluGluHi 114  
Db 30733 TGAATTTGATCGCGTGGCGCGTCAGCGTGGCGTGGTGGTGGCGGTCCAGTGAACG 30674  
Qy 114 sLysHisThr-----GlyGluGlyAr 121  
Db 30673 TGAGCAGACTCTAAACAGATGTTGTTGAGATGGATGTTTTCAGGCAACGAAGGTAT 30614  
Qy 121 gHisLeuGlnThrCysThrAsnSerAspAlaLeuAspProThr-----AlaProAs 139  
Db 30613 TATCGTTATGTCGCGACTAACCGTCCGACGACTTGTATCTCTGCGTGTGTCGTCGAG 30554  
Qy 139 nArgArgAspAsnValAlaPheAlaSerArgArgAspAlaAlaArgArgGluArg 157  
Db 30553 TCGTTTGTACCGTCAGGTTGTTGTTGTTGCCAGAGTACGTGGCGCGCAGAGATTCT 30494  
Qy 158 -----AspGlyThrGlyThrValCysGlnIleThrAs 168  
Db 30493 GAAAGTTTCATATGCGCGGTTCTCTGATACAGATGTTGTATGCTCTCTGTCATTCGTCG 30434  
Qy 168 nGlyGlu-----ThrAspLeuAlaThrMetPheHisLysSerLeuProHi 183  
Db 30433 CGGTACTCCGGGATTCCTGCTGCTGATTTGGCTTAACCTTGTGAATGAAGCTGTTGTT 30374  
Qy 183 sAspGluLeuGlyGln-----ValThrAlaAspPheAlaIleLeuGluAspCy 200  
Db 30373 TGCTGCCCGTGGTAACAGACGCTAGTTTCTATGTTGTAATTCGAAAGCGAAGACAA 30314  
Qy 200 sileLeuAsnGly-----AspPheSerIleCysGluAspVal----- 212  
Db 30313 AATCATGATGGCGCTGAACGCGCTCTATGTTGATGACGGAAGAGCAGAAAGATCAAC 30254





```
Qy 190 rAlaAspPheAlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIleCysG1 210
Db 957 GCGCGGT-----GCGGACGGCCAGTTTCAGATTACCT 989
Qy 210 uAspValProAlaGlyAspProAlaGlyArgLeuValAsnProThrAlaAlaPheAlaI1 230
Db 990 GCAACCGCCACAAGCGAC-----GCCAGGCGCTGGAGATCAGCGCGCGAGCGCGC 1043
Qy 230 eAspIleSerGlyProAlaPheSerAlaThrIleProVal-----ProThrLe 248
Db 1044 TGGCAATATCTCGCCAGTTCAGCATCAGCGG-----CCGACGTCGACCAACCCAGATAC 1100
Qy 248 uSerSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyTrpMetAlaLeuAlaAsAs 268
Db 1101 CACAGCCCTTGAAACACGACCGACCTG-----GCCCTGGCCGATGG 1142
Qy 268 pValProPheMetGlnTyrglyThrAspGluIleThr-----ThrAlaAlaAl 285
Db 1143 CGTCACCTTCACCGGACGCGGTGAGCTGTGTACCGTCAGGTGCGGATGCTGCCGG 1202
Qy 285 aAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSerIleGlySerAspG1 305
Db 1203 TAACCTGATTGGCACCGGT-----GTCTCGGTGCGGATGG 1238
Qy 305 yThrValAspProPheSerGlnLeuPheArgAlaThrPheValGlyValGluThrGlyPr 325
Db 1239 C-----CTGTTACGCTCACCTCTGCGCTGCCAGACC----- 1272
Qy 325 oPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIleThrValGluProLy 345
Db 1273 -----AATGTGAAGCACTCGACGTGCGTCTGTGTGGATGCCG 1310
Qy 345 sGlnGluThrPheAlaPro-----AspLeuAsnTyrrMetValAspPheAspGluTr 362
Db 1311 CGGCAACAGCTCTGCCCATTTGCAATTCGATGCCCGGACATCACACCGCTGACGCGAGT 1370
Qy 362 pLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluGluLeuAr 382
Db 1371 AACCAATATTACAGTTGGTGTGACGG-----CTGGGCCCTGAGCGGTGCGGCGAGCCAGTG 1429
Qy 382 gPheIleArgAsnAlaArg-----AspLeuAlaArgValSerPheValAs 397
Db 1430 CCACTGTGGAAGTGGCGGATGCCAATGCACTGTGTGTCG- GCACGGGTGTCTGGGTGCC 1488
Qy 397 pAsnIleAsnThrGluAlaTyArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSe 417
Db 1489 -AAC-----GGCACCTTCTGTATCGACTCGCCCCCAGCCGC 1523
Qy 417 rArgProGlyIleAsnGlyProPheIleAspSerAspArgGlnAlaGlyPheValAsnPh 437
Db 1524 ACAGCCGGCGGAGCAGTTGAGCTGTGTGTCAGACCGAC-----CCAAGCGGTAATGCTTCAGT 1580
Qy 437 eGlyThrSerHisTyrglyPheArgLeuIleGlyAlaAlaGlu----- 450
Db 1581 GGCTACCGAGTACGATGTACCACTGACCCAGCGCGCGACAGCCCGGAGCAATCTGGCCAT 1640
Qy 451 -----LeuAlaGlnArgAlaSerCysTyrglnIlystrpGlnValHi 464
Db 1641 CGATGCTGATGGCACCACTTCCAGTCCCGGTCACCGGTCGCGCGCTGGCAGCGGTGGARGTGCA 1700
Qy 464 sArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGlyAspLe 484
Db 1701 T-----GAYGCCAACGGCACGTTGATTGGCAGTGCATTTGCCAACGC 1742
Qy 484 uAspAlaAspPheAspIleSerLeu-----LeuGluAsnAspGluLeuLeuIlyAr 501
Db 1743 CGATGGCAGCTTCAGCATCGAACTCGAACTGAATCCCGCACAAAGCCAAACGGCAACTGCTGGACGT 1802
Qy 501 gValala----- 503
Db 1803 GGTGGCCATCATGACGCGCGCGTTCCTCGCTGCCCGCACAGATCATCCGCGCCCGACAT 1862
Qy 504 -----GluIleAsnAl 507
```

```
Db 1863 CACCGCCCTCGCGCACCGACCACTGGCGGTGAGCGCGGACCGGACGCGTTATCACCGG 1922
Qy 507 aAlaGlnAsnProAsnAsnGluValThrTyrglyLeuLeuProGlnAlaIleGlnValGlySe 527
Db 1923 CCGTGGCGAACCGGCGACCACTGGCGCATCGTGGCTGCGAGCGGAACCGAACTGGGCAC 1982
Qy 527 r-----ProThr-----HisProSerTyrglyProSerG1 536
Db 1983 TGGCTGTGTCGGCGCGACCGCGCTGTTACGCTCACTCAACCCCGCCACAGTTGACGG 2042
Qy 536 yHis-----AlaThrGlnAsnGlyAlaPheAlaThrValLeuIlyAlaLe 551
Db 2043 CGAAGTGTCTCAAGCCACCGCGCGATGCGGTGGCAACACTTCCGTTACCAGCTCGGT 2102
Qy 551 uIleGlyLeuAsp---ArgGlyGlyGluCysPheProAsnProValPheProSerAspAs 570
Db 2103 CACCGCGCGGATATCGATGGCGCGATACACCCCG---CCTGAAGCACCGACCAAT-- 2157
Qy 570 pGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyrglyGluIleAsnLy 590
Db 2158 -----CTGCTGATCGCTTCCGCGCACGCTCAGTGTGCGAGCGAA----- 2202
Qy 590 sLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleHisTyrglyPheAspG1 610
Db 2202 ----- 2202
Qy 610 yIleGlnGlyLeuLeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMe 630
Db 2203 -----GCCGCGCATCTACCGTGCAGGTGCGTGACGCGCGCGCGCAACATC-- 2244
Qy 630 tThrPheAlaGluGluAlaThrPheGluPheArgLeuPheThrGlyGluValIleLysLe 650
Db 2245 -----CTCGCGACCGGTACCGTCCGCC----- 2265
Qy 650 uPheGlnAspGlyThrPheSerIle 658
Db 2266 ----GCGGATGGCACCTTCACCGTC 2286
```

Search completed: September 19, 2004, 04:26:20  
Job time : 1105.36 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 19, 2004, 03:58:58 ; Search time 8460.83 Seconds  
(without alignments)  
3463.005 Million cell updates/sec

Title: US-10-691-383-2  
Perfect score: 3528  
Sequence: 1 MLCRAADTRGSPMPDTGVL.....SIDGDMCSGLVYTGVDACQA 676

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US10691383/runat.17092004.102656.1741/app query.fasta\_1.1230  
-DB=GenEmbl -QPMF=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10691383 @CGN 1.1 6663 @runat.17092004.102656.1741 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*

29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3528	100.0	2931	6	AR152190 Sequence
2	3528	100.0	2931	6	BD235853 Gene-modi
3	3528	100.0	2931	6	AR435585 Sequence
4	3528	100.0	2931	8	AF053411 Fucus dis
5	1135.5	32.2	3379	8	LDI491786 Laminaria
6	1129.5	32.0	3420	8	LDI491787 Laminaria
7	626	17.7	1797	6	AX201842 Sequence
8	626	17.7	1908	8	AF218810 Corallina
9	626	17.7	2035	8	BD7657 Corallina p
10	623	17.7	1794	6	E17199 Corallina p
11	617.5	17.5	2029	8	D87658 Corallina p
12	614.5	17.4	1791	6	E17200 Corallina p
13	135.5	3.8	293350	1	EX294153 Pirellula
14	132	3.7	259950	1	EX294136 Pirellula
15	131	3.7	9579	6	AX598603 Sequence
16	131	3.7	11018	1	AE011953 Xanthomon
17	131	3.7	52101	6	AX598593 Sequence
18	131	3.7	340750	1	AX294135 Pirellula
19	130.5	3.7	297850	1	AP006577 Gloeobact
20	130.5	3.7	302550	1	EX294137 Pirellula
21	130	3.7	1499	8	AK110829 Oryza sat
22	129.5	3.7	11679	6	AR318827 Sequence
23	129.5	3.7	16069	1	AE005097 Halobacte
24	129	3.7	347800	1	AP000060 Aeropyrum
25	128	3.6	12639	1	AE005072 Halobacte
26	128	3.6	217594	1	AP002086 Agrobacte
27	126.5	3.6	13522	1	AE007219 Sinorhizo
28	125.5	3.6	299650	1	AP005955 Bradyrhiz
29	124.5	3.5	7185	6	AX803764 Sequence
30	124.5	3.5	11546	1	AE009336 Agrobacte
31	124.5	3.5	13971	1	AE008279 Agrobacte
32	124.5	3.5	61944	6	AX803750 Sequence
33	124	3.5	300861	1	AE016777 Pseudomon
34	123.5	3.5	10910	6	AX024276 Sequence
35	123.5	3.5	10945	6	BD262982 DNA seque
36	123.5	3.5	68750	1	AF210843 Sorangium
37	123.5	3.5	68750	6	AR193029 Sequence
38	123.5	3.5	68750	6	AR199551 Sequence
39	123.5	3.5	68750	6	AR199559 Sequence
40	123.5	3.5	68750	6	AR199567 Sequence
41	123.5	3.5	68750	6	AR201097 Sequence
42	123.5	3.5	68750	6	AR208671 Sequence
43	123.5	3.5	68750	6	BD218234 Genes for
44	123.5	3.5	303450	1	AF005085 Vibrio pa
45	123	3.5	5820	6	AX616791 Sequence

ALIGNMENTS

AR152190	AR152190	2931 bp	DNA	linear	PAT 08-AUG-2001
LOCUS	Sequence 1 from patent US 6232457.				
DEFINITION	AR152190				
ACCESSION	AR152190				
VERSION	AR152190.1	GI:15118240			
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 2931)				
AUTHORS	Vreeland, V. and Ng, K. L.				
TITLE	Recombinant vanadium haloperoxidases and their uses				
JOURNAL	Patent: US 6232457-A 1 15-MAY-2001;				
FEATURES	Location/Qualifiers				
source	1..2931				
	/organism="unknown"				
	/mol_type="unassigned DNA"				
ORIGIN					
Alignment Scores:					
Pred. No.:	2,49e-249	Length:	2931		
Score:	3528.00	Matches:	676		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		
Query Match:	100.00%	Indels:	0		
DB:	6	Gaps:	0		
US-10-691-383-2 (1-676) x AR152190 (1-2931)					
Qy	1	MetLeuCysHisAlaAlaPheAlaThrArgGlySerProMetProAspThrGlyValLeu	20		
Db	228	ATGCTTTGCCATGACGGGACGACGACGACGAGGCTCTCTATGCTGACACCGGAGTCTT	287		
Qy	21	ArgLeuLeuThrSerGluGlnArgAlaLysGlyTyrPargArgGlnLeuGluGlyGluLys	40		
Db	288	CGGTTGTCTCATCAGACGACGCGCTAAAGGTTGGAGACGCGCAGTTAGAGGGGGAGAA	347		
Qy	41	SerLeuGlyPheHisProSerGluThrProTyrIleLysTyrLeuGluGlySerGluThr	60		
Db	348	TCACTAGGTTTTCATCAACGCGAGACGCTTATATCAAGTACTTGGAAAGGCTCTGAGACT	407		
Qy	61	TrpLysLysValLysLeuProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIle	80		
Db	408	TGGAGAGGTTAAGCTTCCACGCGGCGCATATCGGCTTCCAGATCCTGGGTAAATTT	467		
Qy	81	MetAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAla	100		
Db	468	ATGGCCAGGGTCGCGCATCGCTACCGCTTGGCTGTGGTACTGGCGCGCACCTGTTGGCA	527		
Qy	101	PheAspGluValThrAlaSerGlyValPheProGluGluHisLysThrGlyGluGly	120		
Db	528	TTGACAGAGGTACAGCCAGTGGTGTGTTTCCCTGAGGACACAGACACCGGGAGGGA	587		
Qy	121	ArgHisLeuGlnThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArg	140		
Db	588	AGACACCTCCAGACCTGTACAACTCCGACGATGGCTGGATCCGACGGCGCGGAATAGA	647		
Qy	141	ArgAspAsnValAlaPheAlaSerArgArgAspAlaAlaArgArgGluArgAspGlyThr	160		
Db	648	AGGACACAACGTAGCTTTTGGCTCGCGCGGATGCCGCCAGGCGAGAACGTGACGGGACA	707		
Qy	161	GlyThrValCysGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSer	180		
Db	708	GGGACTGCTGTCGCAATACATTAACGAGAACTGATTTGGCTACCATGTTCCCAAGTCT	767		
Qy	181	LeuProHisAspGluLeuGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCys	200		
Db	768	CTGCCACACCATGAATCGGACAGGTAACCGACGACTTCGCTATCCTCGAGGACTGC	827		
Qy	201	IleLeuAsnGlyAspPheSerIleCysGluAspValProAlaGlyAspProAlaGlyArg	220		
Db	828	ATCTTAAACGGAGATTTTCAGCATTTGCGAGGACGTGCTCGGGAGACCGCGGGGTGCG	887		
Qy	221	LeuValAsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThr	240		
Db	888	CTCGTCAATCTTACCGCTGGTTTGCATCGACATATCCGCTCCGCAATCTCGGCTACG	947		
Qy	241	ThrIleProProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeu	260		
Db	948	ACAATACCCCGGTACTACCTTCTCTCTGAGCTGCGCGCTCAGTTGGCGGAGCTA	1007		
Qy	261	TyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGluIleThr	280		
Db	1008	TACTGGATGCGCTGGCCAGGATGTACCTTTTATCAGTATGCGACCGACGAAATATACC	1067		
Qy	281	ThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSer	300		
Db	1068	ACTACCGGCGACCAACCTCGCTGGATGGAGGCTTCCCAATCTGGACGCGGTGTCG	1127		
Qy	301	IleGlySerAspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGly	320		
Db	1128	ATAGGGTCCGATGCTACGGTGGACCGTCTCTCCAGCTCTTCCGAGCGACCTTCGTGGT	1187		
Qy	321	ValGluThrGlyProPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIle	340		
Db	1188	GTTGAAACCGGGCCCTTTGTCTCTCAGCTGCTCGTGACAGCTTCACCATCGACGCTATT	1247		
Qy	341	ThrValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPheAsp	360		
Db	1248	ACGGTCGACCGAAGCAGGAGACATTCGCCCGCGACTTGAACCTATATGTCGATTTTGAC	1307		
Qy	361	GluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluGlu	380		
Db	1308	GAATGCTGAACATTGAGATGTTGGACCCCGCGCGGCCCGGAGAGTTAGACGAAGAG	1367		
Qy	381	LeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsn	400		
Db	1368	CTCGCTTTTATCGGTAAACGCGCGACCTGGCCAGGGTCTCTCTCGTGGACAATATCAAC	1427		
Qy	401	ThrGluAlaTyrArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGly	420		
Db	1428	ACCGAAGCTTATCGCGGGTCTCTTATCTTACTTGTAGCTGGAGGCTTCAGCAGGCCGCT	1487		
Qy	421	IleAsnGlyProPheIleAspSerAspArgGlnAlaGlyPheValAsnPheGlyThrSer	440		
Db	1488	ATCAACGGTCCATTCATCGACAGTGATCGCAGCGGGCTTCGTCAACTTCGCGACGCTCT	1547		
Qy	441	HisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLys	460		
Db	1548	CACCTACTCAGATGTAGTGTGCGCGAGCTGGCGAGCGTGGCTGCTGTGTACCAAAAG	1607		
Qy	461	TrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIle	480		
Db	1608	TGGCAGGTGCATCGATTTGCACGCGCGAGGCTCTCGGGGGTACCTCCACACACCATC	1667		
Qy	481	AlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuAsnAspGluLeuLys	500		
Db	1668	GCGGGGATCTAGATCGACACTTCGACATCTCCCTCTTCTTGAATAATGATGAGCTCTTGA	1727		
Qy	501	ArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuPro	520		
Db	1728	CGTGTGGCGGAGATAATCGCGCGAGAATCCCAACACAGGTACCTACCTTCTTCTCCA	1787		
Qy	521	GlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGln	540		
Db	1788	CAAGCTATCAAGTGGGATCGCCACACGACCTTCTTACCCGTCGCGCACGCTACCCAA	1847		
Qy	541	AsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyCys	560		
Db	1848	AATGGAGCATTTGCCAGTTCCTGAAGGCCCTCATTTGGCTTAGATCGGGAGGTGAGTGC	1907		
Qy	561	PheProAsnProValPheProSerAspGlyLeuLeuIleAsnPheGluGlyAla	580		
Db	1908	TTCCCTAAACCCGCTGTTCACAGCGATGACGGCTGGAACTAATCACTTCGAAAGGGCA	1967		
Qy	581	CysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGln	600		

```

Db 1968 TGCCTTACATATGAGGGAGAGATCAACAAGCTCCGGTCAACGTCGATTTGGAGGAGCAG 2027
Qy 601 MetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuLeuGlyGluThrIle 620
Db 2028 ATGCTGGGCATCCATATCGGTTTCGACGGTATCCAGGCCTACTCTCCGGAGAGACATC 2087
Qy 621 ThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluGluAlaThrPheGluPhe 640
Db 2088 ACTGTAGCAACACTTACCAGGAGCTGATGAGCTTCGCCGAGGAAGCACCTTTGAATTC 2147
Qy 641 ArgLeuPheThrGlyGluValIleIleLeuPheGlnAspGlyThrPheSerIleAspGly 660
Db 2148 CGCTTATTCACCGGAGAGTTCATCAAACTTTTCAGACGGGACATTCCTCATCGATGGA 2207
Qy 661 AspMetCysSerGlyLeuValTyrThrGlyValAlaAlaAspCysGlnAla 676
Db 2208 GATATGTGTTCCGGTTGGTTTACCTGGCTGGCGGACTGCCAGGCT 2255

RESULT 2
LOCUS BD235853 2931 bp DNA linear PAT 17-JUL-2003
DEFINITION Gene-modified vanadium haloperoxidase and utilization thereof.
ACCESSION BD235853
VERSION BD235853.1 GI:33045623
KEYWORDS JP 2002525046-A/1.
SOURCE Fucus gardneri
ORGANISM Fucus gardneri
REFERENCE 1 (bases 1 to 2931)
AUTHORS Eukaryota; stramenopiles; Phaeophyceae; Fucales; Fucaceae; Fucus.
Vreeland,V. and Ng,K.L.
TITLE Gene-modified vanadium haloperoxidase and utilization thereof
JOURNAL Patent: JP 2002525046-A 1 13-AUG-2002;
THE REAGENTS OF THE UNIVERSITY OF CALIFORNIA
COMMENT OS Fucus gardneri
PN JP 2002525046-A/1
PD 13-AUG-2002
PF 27-AUG-1999 JP 2000570298
PR 10-SEP-1998 US 09/151189
PI VALERIE VREELAND, KWAN L NG
PC C12N15/09,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12N9/02,C12N15/
PC 00,C12N5/00
CC Gene-modified vanadium haloperoxidase and utilization thereof
FH Key Location/Qualifiers
FT CDS Location/Qualifiers
FEATURES
source 1..2931
/organism="Fucus gardneri"
/mol_type="genomic DNA"
/db_xref="taxon:3013"

ORIGIN
Alignment Scores:
Pred. No.: 2,49e-249 Length: 2931
Score: 3528.00 Matches: 676
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-691-383-2 (1-676) x BD235853 (1-2931)

Qy 1 MetLeuCysHisAlaAlaAspThrThrArgGlySerProMetProAspThrGlyValLeu 20
Db 228 ATGCTTTGCCATGCGCGGACACACAGAGGCTCTCTATGCTCTGACCGGAGTGCTT 287
Qy 21 ArgLeuLeuThrSerGluGlnArgAlaIleGlySerGlyTrpArgArgGlnLeuGluGlyLys 40
Db 288 CGGTGTGCTCATCAGAGCAGCGCGCTAAAGTTGGAGACGCCAGTTAGAGGGGAGAAA 347
Qy 41 SerLeuGlyPheHisProSerGluThrProTyrIleIleYsrLeuGluGlySerGluThr 60
Db 348 TCACTAGGTTTTCATCCAAAGCGAGACGCTTATATCAAGTACTTTGGAAGGCTCTGAGACT 407

```

```

Qy 61 TrpIysLysValIysLeuProThrAspGlyIleSerAlaSerIleLeuGlyLysIle 80
Db 408 TGAAGAAGGTAAAGCTTCCAAACGGACGGCATATCGGCTTCCAAAGATCTCGGTAAATTT 467
Qy 81 MetAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAla 100
Db 468 ATGCCCAAGGTCGCGATCGCTACCGCTTGGCTGTGTACTGGCCGCGACCTCTGTTGGCA 527
Qy 101 PheAspGluValThrAlaSerGlyValPheProGluGlnHisIleHisThrGlyGluGly 120
Db 528 TTCCGACGAGGTTCACAGCGAGTGGTGTGTTTCCCTGAGGACACACAGACACACCGGGAGGA 587
Qy 121 ArgHisLeuGlnThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArg 140
Db 588 AGACACCTCCAGACCTGTCAAACTCCGACGATCGCTGGATCCGACGGCCGCAATAGA 647
Qy 141 ArgAspAsnValAlaPheAlaSerArgArgAspAlaAlaArgGluArgAspGlyThr 160
Db 648 AGGACCAACGTAGCTTTTGGCTCGCGCGCGATGCCCGCAGGCGAGAACGTCACGGGACA 707
Qy 161 GlyThrValCysGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPheHisIleSer 180
Db 708 GGGACTGTCTGCCAAATCACTAACGGAGAAACTGATTTGGCTACCATGTTCCACAGTCT 767
Qy 181 LeuProHisAspGluLeuGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCys 200
Db 768 CTGCCACACGATGAACCTGGGACAGGTAACCGACAGACTTCGCTATCTTCGAGGACTGC 827
Qy 201 IleLeuAsnGlyAspPheSerIleCysGluAspValProAlaGlyAspProAlaGlyArg 220
Db 828 ATCTTAAACGGAGATTTTTCAGCATTTGCGAGACGTGCTGCGGGAGACCCCGCGGTGCG 887
Qy 221 LeuValAsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThr 240
Db 888 CTGTCATCTACCGCTGCTTGGCTTCCCATGACATATCCGCTCCGCTATCTCGGTACG 947
Qy 241 ThrIleProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeu 260
Db 948 ACAATACCCCGGTACCTACCTCTCTCTGAGCTCGCGCTCAGTTGGCGGAGCTA 1007
Qy 261 TyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGluIleThr 280
Db 1008 TACTGGATGGCGTGGCGAGGATGTACCTTTATGCAGTATGGCAGCGACGAAATTTACC 1067
Qy 281 ThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSer 300
Db 1068 ACTACCGCGGACCAACCTCGCTGGAAATGGAGGCTTCCAAATCTGGACCGCGTGTG 1127
Qy 301 IleGlySerAspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGly 320
Db 1128 ATAGGTCGATGATGATCGGTGGACCGCTTCTCCAGCTTCTCCGAGCGACCTTCGTTGGT 1187
Qy 321 ValGluThrGlyProPheValSerGlnLeuValAsnSerPheThrIleAspAlaIle 340
Db 1188 GTTGAACCGGGCGCTTTGTCTCTCAGCTGCTGCTGAGAACAGCTTCCACCATCGACGCTATT 1247
Qy 341 ThrValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPheAsp 360
Db 1248 ACGTCAACCGAAGCAGGAGACATTCGCCCGGCTTGAACCTATATGTTGCTGATTTGAC 1307
Qy 361 GluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluGlu 380
Db 1308 GAATGGCTGAACATTCAGATGTTGGACCCCGCGCGCGCGAGAGTTAGACGAAGAG 1367
Qy 381 LeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsn 400
Db 1368 CTGCGTTTATTCGGTAAACCGCGCGAGCTGCGCAGGGGTCTCTCTTCGTGGCAATATCAAC 1427
Qy 401 ThrGluAlaTyrArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGly 420
Db 1428 ACCGAAGCTTATCGCGGTCTCTTTATCTTACTTGGAGGCTTTCAGCAGGCCCGGT 1487

```

```
Qy 421 ILeAsnGlyProPheIleAspSerAspArgGlnAlaGlyPheValAsnPheGlyThrSer 440
Db 1488 ATCAACGGTCCATTCATCGACAGTATCGGACGGGGCTTCGTCACATCTTCGGACAGCTCT 1547
Qy 441 HisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLys 460
Db 1548 CACTACTTCAGATTGATAGTGGTCCGCGAGCTGGCGAGCTGGCTCGTGTACCAAAAG 1607
Qy 461 TrpGlnValHisArgPheAlaArgProGluAlaLeuGlyThrLeuHisAsnThrIle 480
Db 1608 TGGCAGGTGCATCGATTGTCACGCCCGGAGCTCTCGGGGGTACCTCCCAACACACATC 1667
Qy 481 AlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLys 500
Db 1668 GCGGGGATCTAGATGCGAGCTTCGACATCTCCCTTCTTGAAATGATGAGCTCTTGAAA 1727
Qy 501 ArgValAlaGluIleAsnAlaAlaGlnAsnProAsnGluValThrTyrLeuLeuPro 520
Db 1728 CGTGTGGCGAGATAAATGCGCGCAGAAATCCCAACACAGAGTCCACTACCTTCTTCCA 1787
Qy 521 GlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGln 540
Db 1788 CAAGCTATCCAAAGTGGATCGCCAAACGACCCCTTCTTACCGTCCGCCACGCTTACCCAA 1847
Qy 541 AsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCys 560
Db 1848 ATGGAGCATTTGCCACAGTTCTGAGGCCCTCATTCGCTAGATCGGGAGGTGAGTGC 1907
Qy 561 PheProAsnProValPheProSerAspAspGlyLeuGluLeuIleAsnPheGluGlyAla 580
Db 1908 TTCCCTTAACCCCGTGTCCCAAGCGATGACGGCTCGAACTAATCAATTCGAAGGGCA 1967
Qy 581 CysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGln 600
Db 1968 TGCCTTACATATGAGGAGAGATCAACAAGCTCGCGTCAACGTCGATTTGGAGGACG 2027
Qy 601 MetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThrIle 620
Db 2028 ATGCTGGGCATCCACTATCGGTTCCGATCCAGGCTTCTTCGAGAGACAAATC 2087
Qy 621 ThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluAlaThrPheGluPhe 640
Db 2088 ACTGTACGAACACTTACCAGGAGCTGATGAGTTCGCCGAGGAAGCCACTTTGAATTC 2147
Qy 641 ArgLeuPheThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIleAspGly 660
Db 2148 CGCTTATTCACCGAGAGGTATCAAACTTTTCCAGACGGGACATTTCTCCATCGATGGA 2207
Qy 661 AspMetCysSerGlyLeuValTyrThrGlyValAlaAspCysGlnAla 676
Db 2208 GATATGTGTTCCGGTTTGGTTTACACTGGCGTGGCGAGTGCAGGCT 2255

RESULT 3
AR435585
LOCUS AR435585 2931 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 1 from patent US 6656715.
ACCESSION AR435585
VERSION AR435585.1 GI:40198553
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 2931)
TITLES Vreeland,V.
JOURNAL Recombinant minimal catalytic vanadium haloperoxidases and their
FEATURES uses
source Patent: US 6656715-A 1 02-DEC-2003;
Location/Qualifiers
1..2931
/mol_type="genomic DNA"
```

```
Alignment Scores:
Pred. No.: 2,496-249 Length: 2931
Score: 3528.00 Matches: 676
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-691-383-2 (1-676) x AR435585 (1-2931)
Qy 1 MetLeuCysHisAlaAlaAspThrThrArgGlySerProMetProAspThrGlyValLeu 20
Db 228 ATGCTTTGCCATCGACGGACACGACAAGAGGCTCTCTATGCTGACACCGGAGTGCCT 287
Qy 21 ArgLeuLeuThrSerGluGlnArgAlaLysGlyTrpArgArgGlnLeuGluGlyLys 40
Db 288 CGGTGTGCTCATCAGACGACGCGCTAAAGGTTGGAGACGCCAGTTAGAGGGGGAGAAA 347
Qy 41 SerLeuGlyPheHisProSerGluThrProTyrIleLysTyrLeuGluGlySerGluThr 60
Db 348 TCACTAGTTTTTCATCCAGCGAGACGCCCTTATATCAAGTACTTGGAGGCTCTGAGACT 407
Qy 61 TrpLysLysValLysLeuProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIle 80
Db 408 TGAAGAAGGTTAAGCTTCCACCGACGCATATCGGCTTCCAAGATCTCTGGGTAAATTT 467
Qy 81 MetAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAla 100
Db 468 ATGGCAGGGTCCGCATCGCTACCGCTTGGCTGTGGTACTGGCGCACCCCTGTTTGCA 527
Qy 101 PheAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGlyGluGly 120
Db 528 TTGACGAGGTTCACAGCCAGTGTGTTCCTCTGAGGAAACACAAGCACACCGGGGAGGA 587
Qy 121 ArgHisLeuGlnThrCysThrAsnSerAspAspAlaLeuAspProThrAlaProAsnArg 140
Db 588 AGACACCTCCAGACCTGTACAAACTCCGACGATCGCTGGATCCGACGGCGCGCAATAGA 647
Qy 141 ArgAspAsnValAlaPheAlaSerArgArgAspAlaAlaArgGluArgAspGlyThr 160
Db 648 AGGACAACTAGCTTTTGGTTCGGCGCGATCGCCGAGGAGAACCTGACGGGACCA 707
Qy 161 GlyThrValCysGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSer 180
Db 708 GGACTGCTGCCAATCATCTACCGAGAACTGATTTGGCTACCATGTTCCACAAGTCT 767
Qy 181 LeuProHisAspGluLeuGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCys 200
Db 768 CTGCCACAGCATGAACCTGGGACAGGTAAACCGCAGACGACTTCGCTATCTCTCGAGGACTGC 827
Qy 201 IleLeuAsnGlyAspPheSerIleCysGluAspValProAlaGlyAspProAlaGlyArg 220
Db 828 ATCTTAAACCGGAGATTTCAGCATTTTGGAGAGCTGCTCTCGGAGAGACCCCGCGGTCGC 887
Qy 221 LeuValAsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThr 240
Db 888 CTGCTCATCTTACCGCTGCTTGGATCGCATCATATCCGGTCCCGCATTTCTCGGCTACG 947
Qy 241 ThrIleProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeu 260
Db 948 ACAATACCCCGGTACCTACCTTTCTCTCTGAGCTCGCGCTCGCTAGTTGGCGGAGCTA 1007
Qy 261 TyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGluIleThr 280
Db 1008 TACTGGATGGCGTGGCCAGGATGATACCTTTATGAGTATGGCACCGACGAAATATACC 1067
Qy 281 ThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyPheProAsnLeuAspAlaValSer 300
Db 1068 ACTACCGCGGACCAACCTCGCTGGAATGGAGGCTTCCCAATCTGGAGCGCGTGTGC 1127
Qy 301 IleGlySerAspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGly 320
Db 1128 ATAGGGTCCGATGGTACGGTGGACCCGCTTCTCCAGCTCTTCCGAGCGACTTCGTTGGT 1187
```

```
Qy 321 ValGluThrGlyProPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIle 340
Db 1188 GTTGAACCGGGCCCTTTGCTCTCAGCTGCTCGTGAACAGCTTCCACCATCGCGCTATT 1247
Qy 341 ThrValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPheAsp 360
Db 1248 ACGGTCAACCGAAGCAGGAGACATTCGCCCGGACTTGAACATATATGCTCGATTTTGAC 1307
Qy 361 GluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluGlu 380
Db 1308 GAATGGCTGAACATTCAGATGTGGACCCCGCGCGCCCGCCGAGAGTTAGACGAAGAG 1367
Qy 381 LeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsn 400
Db 1368 CTGCGTTTATTCGTAACCGCCGCGACCTCGCCAGGGCTCTCTTCGTGGCAATATCAAC 1427
Qy 401 ThrGluAlaTyrArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGly 420
Db 1428 ACCGAAGCTTATCGCGGCTCTCTTATCTACTTGTAGCTGGAGGCTTTCAGCAGCCCGGT 1487
Qy 421 IleAsnGlyProPheIleAspSerAspArgGlnAlaGlyPheValAsnPheGlyThrSer 440
Db 1488 ATCAACGGTTCATTCAGATCGACATCGGAGCGCGGCTTCGTCAACTTCGGCACGTCT 1547
Qy 441 HisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLys 460
Db 1548 CACTACTTCAGATTGATGAGTGGCGCGAGCTGGCGCAGCGTGCCTCGTGTACCAAAAG 1607
Qy 461 TrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIle 480
Db 1608 TGGCAGGTGATCGATTGTCAGCGCCCGCGAGGCTCTCGGGGGTACCCTTCCACACCATC 1667
Qy 481 AlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLys 500
Db 1668 GCGGGGATCTAGATGCGAGCTTCGACATCTCCCTTCTTGAAATGATGAGCTCTTGAAA 1727
Qy 501 ArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuPro 520
Db 1728 CGTGTGGCGAGATAAATCGCGCGCAGAAATCCCAACACGAGGTACCTACCTTCTTCCA 1787
Qy 521 GlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGln 540
Db 1788 CAAAGCTATCCAAAGTGGATCGCCAAACACCTCTTCTTACCGCTCGGCGCACGCTACCCAA 1847
Qy 541 AsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCys 560
Db 1848 AATGGAGCATTTGCCACAGTTCTGAAGGCCCTCATTTGGCTTAGATCGGGAGGTGAGTGC 1907
Qy 561 PheProAsnProValPheProSerAspAspGlyLeuLeuIleAsnPheGluGlyAla 580
Db 1908 TTCCTTAACCCCGTGTTCCTCAAGCGATGACGGCTGGAACTAATCAACTTCGAAGGGCA 1967
Qy 581 CysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGln 600
Db 1968 TGCCTTACATATGAGGAGAGATCAACAAGCTCGCGTCAACGTCGATTTGGAGGCAG 2027
Qy 601 MetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThrIle 620
Db 2028 ATGCTGGGCATCCACTATCGGTCGAGCGTATCCAGGCGCTACTTCTCGAGAGACAAATC 2087
Qy 621 ThrValArgThrLeuHisGlnLeuMetThrPheAlaGluGluAlaThrPheGluPhe 640
Db 2088 ACTGTACGAACACTTCACAGGAGCTGATGACGTTCCCGCGAGGAAGCCACCTTTGAATTC 2147
Qy 641 ArgLeuPheThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIleAspGly 660
Db 2148 CGCTTATTCACCGAGAGGTATCAAACTTTTCCAGACGGGACATTCCTCATCGATGGA 2207
Qy 661 AspMetCysSerGlyLeuValTyrThrGlyValAlaAspCysGlnAla 676
Db 2208 GATATGTGTTCCGGTTTGGTTTACACTGGCGGTGGCGAGCTGCACGGCT 2255
```

## RESULT 4

AF053411

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

5'UTR

CUS

AF053411 2931 bp mRNA linear PLN 12-SEP-1998  
Fucus distichus vanadium bromoperoxidase mRNA, complete cds.

AF053411.1 GI:3582762

Fucus distichus

Fucus distichus

Eukaryota; stramenopiles; Phaeophyceae; Fucales; Fucaceae; Fucus.

1 (bases 1 to 2931)

Vreeland, V.; Ng, K. and Epstein, L.

Direct Submission

Submitted (11-MAR-1998) Environmental Science, Policy and

Management, University of California, 201 Wellman Hall, Berkeley,

CA 94720-3112, USA

Location/Qualifiers

1..2931

/organism="Fucus distichus"

/mol\_type="mRNA"

/db\_xref="taxon:3012"

/dev\_stage="2-cell embryo"

1..227

228..2258

/codon\_start=1

/product="vanadium bromoperoxidase"

/protein\_id="AAC35279.1"

/db\_xref="GI:3582763"

/translation="MLCHAADTTGSGMPDPTGVRLTLTSEQRAGKWRQLEGEKSLGP

HPSETYIKYLGSESTWKKVKLPTDGISASKILGKIMARVRIATATAVLAAPCLAFD

EVTSAGVPEEHKHTGEGRHQLQCTNSDDALDPTAPNRDNVAFASRRDARERDGT

GTVCQTNGETDLATMFHKSLSHDLQVTDADDFAILEDCLNGDFSCIDBVPAGDPA

GRLVNPTAAFAIDISGPAFAFSATTIPVPTLSSPELAQAELYMALARDVPFMQYGT

DETTTAAANLAGMGFPNLDVSIIGSDGTVDFPSOLFRAATFVGVETGPPVSOLLVNS

FTDAITVSPKQSTFAPDLNMYVDFDEMLNIONGPPAGPELDEBELRFTRNARDLAR

VSFVDNINTEARNGSLILLEGAFSPRGINGPIDSDRQAGFVNFGTSHYFRILGANE

LAQRASYCQVHRFARPEALGTLHTITAGDLADFDISLEDELLKKRAEINAA

QPNNVTVLLPQAIQVSGTSPHSYSHATQNGAFATVLKALIGLDGCGECPNPVF

PSDGLIELNFEGACLTYSGEINKLVAVFGRQMLGIHYRFDGIQGLLIGETITVRT

LHQELMTFABEATFEFLFTGEVILKFDQGTFSIDGMCSGLVTVTGADCA"

2259..2931

3'UTR

ORIGIN

Alignment Scores:

Pred. No.: 2,49e-249 Length: 2931

Score: 3528.00 Matches: 676

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 8 Gaps: 0

US-10-691-383-2 (1-676) x AF053411 (1-2931)

Qy 1 MetLeuGlyHisAlaAlaAspThrArgGlySerProMetProAspThrGlyValLeu 20

Db 228 ATGCTTTGCCATCGACGGACACGACAAAGAGGCTCTCTATGCTGACACCGAGTGCCT 287

Qy 21 ArgLeuLeuThrSerGluGlnArgAlaLysGlyTyrArgArgGlnLeuGluGlyLys 40

Db 288 CGGTTGCTTCATCAGACGACGCGCTAAAGTTGGAGACCCAGCTTAGAGGGGAGAAA 347

Qy 41 SerLeuGlyPheHisProSerGluThrProTyrIleLysTyrLeuGluGlySerGluThr 60

Db 348 TCACTAGGTTTTCATCCAGCGAGACGCTTATATCAAGTACTTGAAGGCTCTGAGACT 407

Qy 61 TrpLysIleValLysLeuProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIle 80

Db 408 TGAAGAAGGTTAAGCTTCCACGCGGACGATATCGGCTTCCAGATCTCCGGGTAAATTT 467

Qy 81 MetAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAla 100

Db 468 ATGGCAGGGTCCGCAATCGCTACCGCTTGGCTGTGTGTACTGGCCGCCACCTGTTTGCA 527

101 PheAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGlyGluGly 120  
120  
528 TTGGACAGGTCACAGCCAGTGGTGTTCCTTGAGGAAACACAGACACACCGGGGAGGA 587  
121 ArgHisLeuGlnThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArg 140  
140  
588 AGACACCTCCAGACCTGTACAAACTCCGACGATCGCGTGGATCCGACCGCGCGAATAGA 647  
141 ArgAspAsnValAlaPheAlaSerArgArgAspAlaAlaArgGluArgAspGlyThr 160  
160  
648 AGGACAACTAGCTTTGGCTCGCGCGGATGCCGCCAGGAGACGTCGCGGACA 707  
161 GlyThrValCysGlnThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSer 180  
180  
708 GGAAGTGTCTGCCAAATCACTAACCGGAGAACTGATTGGCTACCATGTTCCCAAGTCT 767  
181 LeuProHisAspGluLeuGlyGlnValThrAlaAspAspPheAlaLeuLeuGluAspCys 200  
200  
768 CTGCCACACGATGAACCTGGGACAGGTAAACCGACGACCTTCGCTATCCTCGAGGACTGC 827  
201 IleLeuAsnGlyAspPheSerIleCysGluAspValProAlaGlyAspProAlaGlyArg 220  
220  
828 ATCTTAAACCGGATTTTCAGCATTTGCGAGGACGTCCTCGCGGAGACCCGCGGGTFCG 887  
221 LeuValAsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThr 240  
240  
888 CTGCTCAATCTTACCGTGGTGGTGGCATCGACNATCCGGTCCCGCATTTCTGGCTACG 947  
241 ThrIleProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeu 260  
260  
948 ACAATACCCCGGTACCTACCTTTCCTCTCGAGCTCGCGCTCAGTTGGCGGAGCTA 1007  
261 TyrTrpMetAlaLeuAlaArgAspValProPheMetGlnThrGlyThrAspGluIleThr 280  
280  
1008 TACTGGATGGCGTGGCGGAGTGTACCTTTATGCAGTATGSCACCCGACGAAATPACC 1067  
281 ThrThrAlaAlaLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSer 300  
300  
1068 ACTACCGCGGAGCAACCTTCGCTGGAAATGGAGGCTTCCCAATCTGGACCGCGTGC 1127  
301 IleGlySerAspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGly 320  
320  
1128 ATAGGGTCCGATGATCGGTGGACCCGCTTCCAGCTCTCCGAGCGACCTTCGTGTGT 1187  
321 ValGluThrGlyProPheValSerGlnLeuValAsnSerPheThrIleAspAlaIle 340  
340  
1188 GTTGAACCGGGGCTTTGTCTCTCAGCTCTGCTGAAACAGCTTCACCATCGACGCTATT 1247  
341 ThrValGluProLysGlnThrPheAlaProAspLeuAsnTyrMetValAspPheAsp 360  
360  
1248 ACGGTGCAACCGAAGCAGGAGACATTCGCCCCGACCTTGAACTATATGCTCGATTGTGAC 1307  
361 GluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluGlu 380  
380  
1308 GAATGGCTGAACATTCAAGATGGTGGACCCCGCGCGCCCGAAGAGTTAGACGAAGAG 1367  
381 LeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsn 400  
400  
1368 CTGGCTTTATTCGGTAACCGCGGACCTGCGCCAGGCTCTCTTGGTGGCAATATCAAC 1427  
401 ThrGluAlaTyrArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGly 420  
420  
1428 ACCGACCTTATCGGGGTCTCTTATCTTACTTGGCTGGAGCTTCAGCAGCGCCGGT 1487  
421 IleAsnGlyProPheIleAspSerAspArgGlnAlaGlyPheValAsnPheGlyThrSer 440  
440  
1488 ATCAACGGTCAATTCATCGACAGTATCGGCGGCGCTTCGTCAACTTCGCGCACGTCT 1547  
441 HisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLys 460  
460  
1548 CACTACTTCAGATTGATAGTGGCGCGGAGCTGGCGAGCGGTGCTCGTGTACCAAAAG 1607  
461 TrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIle 480

1608 TGSCAGGTGCATGATTTTCACGCCCGAGGCTCTCGGGGTACCCTCCCAACACCATC 1667  
481 AlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLys 500  
500  
1668 GCGGGGATCTAGATCGACACTTCGACATCTCCCTCTTTGAAATGATGAGCTTTGAAA 1727  
501 ArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuPro 520  
520  
1728 CGTGTGGCGAGATAAATGCGGCGAGATCCCAACACGAGGTCACTTACTTCTTCCA 1787  
521 GlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGln 540  
540  
1788 CAAGCTATCAAGTGGGATCGCCACGACCCCTTCTTACCCGTCGCGCACGCTACCCAA 1847  
541 AsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyCys 560  
560  
1848 AATGGAGCATTTGCCACAGTTCTGAAGGCCCTCATTTGGCCTAGATCGGGAGGTGAGTGC 1907  
561 PheProAsnProValPheProSerAspAspGlyLeuGluLeuIleAsnPheGluGlyAla 580  
580  
1908 TTCCCTTAACCCCGTGTTCACAGCGATGACGGCTGGAACCTAATCACTTCGAAGGGCA 1967  
581 CysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGln 600  
600  
1968 TGCCTTACATATGAGGAGAGATCAACAAGCTCGCGTCAACGCTCGCATTTGGAGGCAG 2027  
601 MetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThrIle 620  
620  
2028 ATGCTGGCATTCACATATCGGTTCGCGGTATCCAAAGGCCCTACTTCTCGAGAGACAATC 2087  
621 ThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluAlaThrPheGluPhe 640  
640  
2088 ACTGTAGAACACTTCCACGAGAGCTGATGACGTTGCCGAGGAGGCCACCTTTGAATTC 2147  
641 ArgLeuPheThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIleAspGly 660  
660  
2148 CGCTTATTCACCGAGAGGTCATCAAACTTTTCCAGGCGGACATTTCTCATCGATGGA 2207  
661 AspMetCysSerGlyLeuValTyrThrGlyValAlaAspCysGlnAla 676  
676  
2208 GATATGTGTTCCGGTTTGGTTTACACTGCGCGGAGACTGCCAGGCT 2255

RESULT 5  
LDI491786 3379 bp mRNA linear PLN 28-JUN-2003  
LOCUS Laminaria digitata mRNA for vanadium-dependent bromoperoxidase 1  
DEFINITION (VBPO1 gene).  
ACCESSION AJ491786  
VERSION AJ491786.1 GI:32329411  
KEYWORDS vanadium-dependent bromoperoxidase 1; VBPO1 gene.  
SOURCE Laminaria digitata  
ORGANISM Laminaria digitata  
Eukaryota; stramenopiles; Phaeophyceae; Laminariales;  
Laminariaceae; Laminaria.  
REFERENCE 1  
AUTHORS Colin,C., Leblanc,C., Wagner,E., Delage,L., Leize-Wagner,E., Van  
Dorsseiler,A., Kloareg,B. and Potin,P.  
TITLE The brown algal kelp Laminaria digitata features distinct  
bromoperoxidase and iodoperoxidase activities  
J. Biol. Chem. 278 (26), 23545-23552 (2003)  
JOURNAL 22703831  
MEDLINE 12697758  
PUBMED 12697758  
REFERENCE 2 (bases 1 to 3379)  
AUTHORS Leblanc,C.  
TITLE Direct Submission  
JOURNAL Submitted (26-JUN-2002) Leblanc C., UMR 1931, CNRS-Goemar, Station  
Biologique, BP 74, 29682 Roscoff, FRANCE  
FEATURES  
source Location/Qualifiers  
1..3379 /organism="Laminaria digitata"  
/mol\_type="mRNA"  
/db\_xref="taxon:80365"





```
QY 582 -----LeuThrTyrrGluGlyGluLeuAsnLeu 591
Db 1608 GTCAACGAGAGGCAAGCTCGTGAAGGCTCAGTACGAGGAGAGCTCAACAGATC 1667

QY 592 AlavalAsnValalaphGlyArgGlnMetLeuGlyIleHisTyrrArgPheAspGlyIle 611
Db 1668 AGCCCAACAGTGTCTTTGGGAAGTCCACATCGTGTTCACCTGAGGATGGACGGCGTG 1727

QY 612 GlnGlyLeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThr 631
Db 1728 TACGAGAGCGCTCATGGAGAGACCACTGCTGCTCGCGCTGACGAGCAACTTCTTGGC 1787

QY 632 PheAlaGlu-----GluAlaThr 637
Db 1788 CTCCCGAGGCGCGAGGTGAGGCAAGAGCGCGTGTGACATCCCCCGGCGACT 1847

QY 638 PheGluPheArgLeuPheThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSer 657
Db 1848 TACAAATTCGCGCTTTACAGCGGTAAAGATTCTTGAGCTCTACGCGAGGAACCTTGTAACAG 1907

QY 658 IleAspGlyAspMetCysSerGlyLeuValTyrrThrGlyValalalAspCys 674
Db 1908 CTGGACGCAAGCTCTGCGAGGGG---GCCTTACCGCGGACGACTTCTGT 1955

RESULT 6
LOCUS LDI491787 3420 bp mRNA linear PLN 28-JUN-2003
DEFINITION Laminaria digitata mRNA for vanadium-dependent bromoperoxidase 2
(VBPO2 gene).
ACCESSION AJ491787.1 GI:32329413
VERSION vanadium-dependent bromoperoxidase 2; VBPO2 gene.
KEYWORDS Laminaria digitata
SOURCE Laminaria digitata
ORGANISM Eukaryota; stramenopiles; Phaeophyceae; Laminariales;
Laminariaceae; Laminaria.
REFERENCE 1
AUTHORS Colin,C., Leblanc,C., Wagner,E., Delage,L., Leize-Wagner,E., Van
Dorselaer,A., Kloareg,B. and Potin,P.
TITLE The brown algal kelp Laminaria digitata features distinct
bromoperoxidase and iodoperoxidase activities
J. Biol. Chem. 278 (26), 23545-23552 (2003)
MEDLINE 27033831
PUBMED 12697758
REFERENCE 2 (bases 1 to 3420)
AUTHORS Leblanc,C.
TITLE Direct Submission
SUBMITTED (26-JUN-2002) Leblanc C., UMR 1931, CNRS-Goemar, Station
Biologique, BP 74, 29682 Roscoff, FRANCE
FEATURES
source
1. .3420
/organism="Laminaria digitata"
/mol_type="mRNA"
/db_xref="taxon:80365"
/tissue_type="sporophyte"
/clone_lib="lambda ZAP II - S7"
/country="France: Brittany"
1. .3420
/gene="VBPO2"
1. .36
/gene="VBPO2"
37. .2085
/gene="VBPO2"
/codon_start=1
/product="vanadium-dependent bromoperoxidase 2"
/protein_id="CA037192.1"
/db_xref="GI:32329414"
/tranlation="MKPAPTOPCALYSVLRLHVACALCVACAFLLFEYQISSLGL
TALLAFAPACGEEPEPTQPLLSNVCVRSLDFDPVPRKVTLLKRLALAKDE
ISVGTFCNVNGDEENPEPTFAGQHTLPHDKFGQVDEDAKLLKLCVFTDINECEK
VPSGARGGAKLTNPVGTAHQVTGADSDNVITTPDSLLSLRLAAQAEVYWMALL
RDIPFGEFAKNDYVRLAAENLQSLPAPKGLNIPRSEGGKIDPVTDLFRITWPVGTTGP
```

```
VVSQFMLSDFLIDSIKVTPKADPLTFCVDYMTAFQWLDVQNGASKLETTFFBENPRF
IRNGRLATIALRDOLYTEAFRAALITFTGALGGVGPVAAEAEQOGFATFGEPHIL
TAMASSSTRHAWYAKWQVHRMLRPEAYGALVHNTLRDVIITPLPDSILRNTELLNR
VEVHNRMPDGEKTELLPMAAAGSPTHPAYPSHAINNGAVITALKAFGLSVEAGOK
CFPNVVSNDGRTKRIKPSGREIVGECVNEKGLVEGTITGELNKNISANVILGRS
HIGVHRMDGVYALMGETSCVRLQOELGFLPEAREVEGKRRGDIIPATYKFLYS
GKILELYGRNLYKLDGLKCEGAFGTDFDIDEDDYSSFFDDIVVEHAQFSLHTEHL
"
2086. .3420
/gene="VBPO2"

3'UTR
ORIGIN
Alignment Scores:
Pred. No.: 3,41e-73 Length: 3420
Score: 1129.50 Matches: 253
Percent Similarity: 51.82% Conservative: 102
Best Local Similarity: 36.93% Mismatches: 247
Query Match: 32.02% Indels: 83
DB: 8 Gaps: 14

US-10-691-383-2 (1-676) x LDI491787 (1-3420)
QY 51 TyrIleLysTyrrLeuGluGlySerGluThrTyrrLysLysValLysLeuProThrAspGly 70
Db 7 CATACCAAGAGTTTAGACACACAGCCTGCCATGAAGCGCTAGGCCACCCAGCCGCGC 66

QY 71 IleSerAlaSerLysIleLeuGlyLysIleMetAlaArgValArgIleAlaThrAlaLeu 90
Db 67 -----GGCGTGTACTGTCTCTCTCTCTCGACACGTTGGGTGTGCCTG 111

QY 91 AlaValValLeuAla----- 95
Db 112 TGGTGTGTGCGATCGCGCTTTCTTTTGTTCGAATATCAGATTCTTCTTGGGCTTGACC 171

QY 96 -----AlaProCysLeuAlaPheAspGluValThrAlaSerGlyVal 109
Db 172 GCTTTGTGCGCGGCTTTCGCTCCTCTCTCTCGCTCGGCTACGAGGAG----- 213

QY 110 PheProGluGluHisLysHisThrGlyGluGlyArgHisLeuGlnThrCysThrAsnSer 129
Db 214 ---CCTCCGAGCGCCAGCGAGCTCTGCTCAGCGGAGACGTGTGCAGGTGCGAGACTCC 270

QY 130 AspAspAlaLeuAspProThrAlaProAsnArgArgAspAsnValAlaPheAlaSerArg 149
Db 271 CTCGACTCTCTGACCGCGTGGCG-----CGCGGAAGGTGACCTGCTC---AAG 318

QY 150 ArgAspAlaAlaArgArgGluArgAspGlyThrGlyThrValCysGlnIleThrAsnGly 169
Db 319 CGCTGCGCCATCGCAAGGACGAGATATCGTTGGACCCCACTTGCACGCTCAACAACGCGC 378

QY 170 GluThrAsp-----LeuAlaThrMetPheHisLysSerLeuProHisAspGlu 185
Db 379 GACGAGGAACGCTCCGCTCTTCGCTGGCGAGTACCACAGAGCTTACCGCAGCAAG 438

QY 186 LeuGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCysIleLeuAsnGlyAsp 205
Db 439 TTCGTCAGGTGGACGAGCAGCATACAAGAAGCTGTGAGTGCCTTTCACGAGCGAC 498

QY 206 PheSerIleCysGluAspValProAlaGlyAspProAlaGlyArg----- 220
Db 499 ATCAACGAGTGCAGAAAGGTGCGCTTCGCG-----GCCGCGCAAGAGGTGGCGCGAAG 552

QY 221 LeuValAsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThr 240
Db 553 CTGACCAACCCCTTCGGGGGCGACCGCGCAGGTACCGGTGCTGACGAGCAATGTG 612

QY 241 ThrIleProProValProThrLeuSerSerProGluLeuAlaGlnLeuAlaGluLeu 260
Db 613 TTCATCACCGCGCGCAGCTCTCTCTCCGAAAGGTGCGCGCGCAGCAGCAAGATT 672

QY 261 TyrrPheMetAlaLeuAlaArgAspValProPheMetGlnTyrrGlyThrAspGluIleThr 280
Db 673 TACTGGATGGCCCTGTGCTTAGGGACATCCCTTTCGGGAGTTCGCGCAAGAAAGACTACGTC 732
```

Qy	281	ThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSer	300
Db	733	AGGCTCGCGCAGAGAACCCTGCAAGACCTGCGCGGTTCAGAGGACTCAAC--ATTCCC	789
Qy	301	IleGlySerAspGlyThrValAspProPheSerClnLeuPheArgAlaThrPheValGly	320
Db	790	CGGAGCAGGAGGAGAAAGATCGACCCCTGTCTCAGTATCTGTTCGGAACGACTGGCCGGGT	849
Qy	321	ValGluThrGlyProPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIle	340
Db	850	GTGNACACCGGTCCCGTCGTCTCAGTTATGCTGTTCGGACTTCCTGTATCGACTCTATC	909
Qy	341	ThrValGluProIleGlnGlnThrPheAlaProAspLeuAsnThrMetValAspPheAsp	360
Db	910	AAGGTACCCCGAAGGCTGACCCCTTACCCAGGGGTGGACTACATGATCGCGCTTCCAG	969
Qy	361	GluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluGlu	380
Db	970	CCGTGGCTTCAGCTACAGACGGGGCTTCCAAGCTTGAACACTCTTCGACGAGGAGAAC	1029
Qy	381	LeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsn	400
Db	1030	CCCCGCTTCATCCGACCGCCGAGACCTGGCCACTATCGCTTGAGGATCAACTCTAC	1089
Qy	401	ThrGluAlaTyrArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGly	420
Db	1090	ACGGAGGCTTTCGCGCGCGGCTCATCTCTGTTCACGGAAGGTGCACTTGGC----	1143
Qy	421	IleAsnGlyProPheIleAspSerAspArgGlnAlaGlyPheValAsnPheGlyThrSer	440
Db	1144	GAGTCCGGGCTTACGCGGAGGCGCGAGCGGCGCAGCAGGGGTTCGCCACTTTCGGGAGCCG	1203
Qy	441	HisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLys	460
Db	1204	CACATCTCACCGCATGGCTCGGCCAGCTTCTCCACGCGCCACGGTGGTACGCTAAG	1263
Qy	461	TrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIle	480
Db	1264	TGGCAAGTCCACCGCATGCTCGGACCGAGGGGTACGGTCCCTGGGTACATACACGCTC	1323
Qy	481	AlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLys	500
Db	1324	ATGAGGACGCTCATCACCCGCTCGCGACTCTATCTCTGAGAACACGGAGCTCTCAAC	1383
Qy	501	ArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuPro	520
Db	1384	CGAGTGGAGGTGCACACCAACCGATGAAACCCAGACGCGGAGAGACCTTCTCTGCTCCC	1443
Qy	521	GlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGln	540
Db	1444	ATGGCGCAGCCACCGGCTCTCCACACGACCCCGCTTACCCAGTGAGACAGCATCAAC	1503
Qy	541	AsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCys	560
Db	1504	AACGGCGCTTACATCACCGCGCTCAAGGCGTTCCTCGGTCACGAGCGCGCCAGAGTGC	1563
Qy	561	PheProAsnProValPheProSerAspAspGlyLeuGluLeuIleAsnPheGlu-----	578
Db	1564	TTCCCCAACCCCGTGTCTCGAACGACGAGGGAACCAAGAGGATCAATAACAAGCCCTCC	1623
Qy	579	-----GlyAlaCys-----LeuThr	583
Db	1624	GGAAGGAGATGTGGCGGAGTGCGTCAACGAGAAGGCGAGCTGCTGCAAGGCTCTCAGC	1683
Qy	584	TyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGly	603
Db	1684	TACGAGGAGAGCTCAACAGATACGAGCCCAACGTGCTCTTGGGAAGGTCCCATCGGT	1743
Qy	604	IleHisTyrArgPheAspGlyIleGlnGlyLeuLeuLeuGlyGluThrIleThrValArg	623
Db	1744	GTTCACTGGAGGATGGACGGCGTGTACGAGCGCTCATGGGAGAGACAGCTGCGTCCGC	1803

Qy	624	ThrlEuHisGlnGluLeuMetThrPheAlaGlu-----	634
Db	1804	CGCTGCAGCAGGAACCTTCTGGCCTCCCGAGGCGCGAGGTGTAGGGCAAGAAGCGC	1863
Qy	635	-----GluAlaThrPheGluPheArgLeuPheThrGlyGluValIleuys	649
Db	1864	CGTGTGCACATCCCGCGGCGACTTACAAATTCGCCCTTTACACGGTAGATTCCTTGAG	1923
Qy	650	LeuPheGlnAspGlyThrPheSerIleAspGlyAspMetCysSerGlyLeuValTyrThr	669
Db	1924	CTCTACGCGAGGAACCTGTGTACAACTGTCGCGCAAGCTCTGCGAGGG---GCCTTCACC	1980
Qy	670	GlyValAlaAspCys	674
Db	1981	GGCGACGACTTCTGT	1995
RESULT 7			
AX201842			
LOCUS	AX201842	1797 bp	DNA linear PAT 30-AUG-2001
DEFINITION	Sequence 1 from Patent WO0153494.		
ACCESSION	AX201842		
VERSION	AX201842.1	GI:153931686	
KEYWORDS	Corallina officinalis		
SOURCE	Corallina officinalis		
ORGANISM	Corallina officinalis		
	Eukaryota; Rhodophyta; Florideophyceae; Corallinales;		
	Corallinaceae; Corallinoideae; Corallina.		
REFERENCE	1		
AUTHORS	Butler, A. and Carter, J.N.		
TITLE	Cloning and overexpression of vanadium bromoperoxidase from marine algae		
JOURNAL	Patent: WO 0153494-A 1 26-JUL-2001;		
	The Regents of the University of California (US)		
FEATURES	Location/Qualifiers		
source	1..1797		
	/organism="Corallina officinalis"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:35170"		
ORIGIN			
Alignment Scores:			
Pred. No.:	1.28e-36	Length:	1797
Score:	626.00	Matches:	176
Percent Similarity:	45.00%	Conservative:	76
Best Local Similarity:	31.43%	Mismatches:	220
Query Match:	17.74%	Gaps:	88
DB:	6		15
US-10-691-383-2 (1-676) x AX201842 (1-1797)			
Qy	168	AsnGlyCyluThrAspLeuAlaThrMetPheHisysSerLeuProHisAspGluLeuGly	187
Db	148	ARCGGAGATCCAAGCTTCATCGTTAGCTTCACAAAGGGTCTTCGCGATGACGACATGGC	207
Qy	188	Gln---ValThrAlaAspPheAlaIleLeuGluAspCysIleLeuAsnGlyAspPhe	206
Db	208	GCTATTATTGACCCCGAGGACTTCTGGCCCTTCGTTCTGGCATCAATAGTGGAGATGAG	267
Qy	207	SerIleCysGluAspVal-----ProAlaGlyAspProAlaGlyArgLeu-----	221
Db	268	AAGGAGATCGCCGACCTCACATTGGGGCCAGCTCGCGACCCGACACTGGCTTACCAATC	327
Qy	222	-----ValAsnProThr	225
Db	328	TGGCGCTCGGATTGGCGAATTCCTCGAGCTCGAAGTCGAGGATCGGAAACAGCTCT	387
Qy	226	AlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThrThrIleProProVal	245
Db	388	GCCGGTCTACCTTCGACCTGGAGGGCCCGAGCGCGCAGTCGATTGCCATGCCACCGGCG	447
Qy	246	ProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeu	265
Db	448	CTTGTGCTCAGAGCCCTGAGCTCATCTCGCGAGATAGCAGAACTGTACTGTATGGCGCTT	507

```
QY 266 AlaArgAspValProPhe-----MetGlnTyr 274
Db 508 GGACGGCAATCGAGTTTACGAGTTGATTCCTCCCAAGAACGACGAGGTATATTCAGTTT 567
QY 275 GlyThrAspGluIleThrThrAlaAlaAsnLeuAlaGlyMetGlyGlyPhePro 294
Db 568 GCTATTGATCAGCTTAACGGGCTAGAGTGGTTCAACACGCGCGCAATGCTCGGAGATCCG 627
QY 295 AsnLeuAspAlaValSerIleGlySerAspGlyThrValAspProPheSerGlnLeuPhe 314
Db 628 CCTCGCGAATTCGTCGCCCTCGCGGTGAGTGACTGTC-----GGAACCTTGTTTC 678
QY 315 ArgAlaThrPheValGlyValGluThrGlyProPheValSerGlnLeuLeuVal----- 332
Db 679 CGCGGTATTCTCCAGGCTCTGAGGTGCGCCCGTATCTCAGCCAGTACATCATCGTTGGT 738
QY 332 ----- 332
Db 739 AGCAAGCAGATTGGCTCCGCGCGGGTGCGCAACAAACTCTCGTACGCCCCAATGCCGCT 798
QY 333 -----AenSerPheThrIleAspAlaIleThrValGluProLysGlnGluThr 348
Db 799 GATGAGTTTATGATCGTAATCGCCTACGGAAGCATCACCATTAGCCACGCGGTCCGTATC 858
QY 349 PheAlaProAspLeuAsnTyrMetValAspPheAspGluTrpLeuAsnIleGlnAsnGly 368
Db 859 GCCACGCTCGGGCGGACTTTATGACCGATTGAAGGTATTCCTTGACGTCGACGAGCT 918
QY 369 GlyProAlaGlyProGluGluLeuAspGluGluLeuArgPheIleArgAsnAlaArg 388
Db 919 CGCGACTTCGAGGCTTTGATCGTATGAGCGCGGAGCAGCGCTCATCCGACGATCCGC 978
QY 389 AspLeuAlaArgValSerPheValAspAsnIleAsnThrGluAlaTyrArgGlySerLeu 408
Db 979 GATCTTCGCGAGTGGGTGACCTTTGACGCACTGTACGAGGCTACCTCAATGCGTGCCTG 1038
QY 409 IleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSer 428
Db 1039 ATTCTGTGGCGAACCGCGTGGCTGCGTTCGATCCCAACATT-----CGTTCCAGCAGGAG 1092
QY 429 AspArg-----GlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeu 445
Db 1093 GACAAGCTCGATAACACGAGGACGCTTTGTGAACCTCGGAGACGACACGCTGAGTCTG 1152
QY 446 Ile---GlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnValHis 464
Db 1153 GTGACTGAGGTGGCTACGCGCGGCTTGAGCGCGTACGCTACGAGAGTTTAAACATTAT 1212
QY 465 ArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHis---AsnThrIleAlaGlyAsp 483
Db 1213 CGTCGCTCGGCTGAGGCTACCGGTGGTGTGATAGCTGAGTGAACAAATCGCAGCGGAG 1272
QY 484 -----LeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeu 499
Db 1273 AAGGGCAGAGCGCTTTCCCTGAGTTGACCTTCTGTTGAAGAGCTTGAAGATATCTCTG 1332
QY 500 LysArgValAlaGluIleAsnAlaAlaGlnAsn-----ProAsnAsnGlu 514
Db 1333 GAGAAAGCTGAAATTAGCAATAGGAAACAGAAATCTGCTGACGAGAGATCTCGACCTGAT 1392
QY 515 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 534
Db 1393 CCTTCATCTCTGTTCGCGAGGCAATTCGCGAGGCGAGCCCATCTCCATCCGCTCTACCGA 1452
QY 535 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGly--- 553
Db 1453 AGCGGCCACGCTGTGGTGTGCTGGCGCATGTGTGACGATCTCAAGGGCTTCTTCGACTCC 1512
QY 554 -----LeuAspArgGlyGlyGluCysPheProAsnProValPheProSerAsp 570
Db 1513 AACTTCAGATCGATCAGGTGTTTCGAG-----GTGCAAAAGATGAGGAC 1557
```

```
571 GlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnLys 590
1558 AACCTTGTAAAGTCGCTTTTCAAGGGACT---CTCACTGTTCCGCGTGAAGTGAACAAG 1614
QY 591 LeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGly 610
Db 1615 CTGCGCGCAATATTGCGATCGCGCGGAACATGCGAGGTGTTCTACTACTTCTCTGACCAG 1674
QY 611 IleGlnGlyLeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMet 630
Db 1675 TTCCAGTCAATTCCTGTTGTTGAGCAGGTTCGCAATTCGGAATCTTGAAGACGAGTCTG 1734
QY 631 ThrPheAlaGluGluAlaThrPheGluPheArgLeuPheThrGlyGluValIleLysLeu 650
Db 1735 ACATATGCGGAGAACTTCTTCACTTCCCACTGCGCAAGTTTGTGGAACATCAATCCAGATC 1794

RESULT 8
AF218810 1908 bp mRNA linear PLN 14-AUG-2002
LOCUS Corallina officinalis vanadium-dependent bromoperoxidase mRNA,
DEFINITION complete cds.
ACCESSION AF218810 GI:21309839
VERSION AF218810.1
KEYWORDS
SOURCE Corallina officinalis
ORGANISM Corallina officinalis
Eukaryota; Rhodophyta; Florideophyceae; Corallinales;
Corallinaceae; Corallinoideae; Corallina.
REFERENCE 1 (bases 1 to 1908)
AUTHORS Carter, J.N., Beatty, K.E., Simpson, M.T. and Butler, A.
TITLE Reactivity of recombinant and mutant vanadium bromoperoxidase from
the red alga Corallina officinalis
JOURNAL J. Inorg. Biochem. 91 (1), 59-69 (2002)
MEDLINE 22117781
PUBMED 12121762
REFERENCE 2 (bases 1 to 1908)
AUTHORS Carter, J.N. and Butler, A.
TITLE Direct Submission
JOURNAL Submitted (23-DEC-1999) Chemistry, University California Santa
Barbara, Santa Barbara, CA 93106, USA
FEATURES
source
1..1908
/organism="Corallina officinalis"
/mol_type="mRNA"
/db_xref="taxon:35170"
1..1197
/codon_start=1
/product="vanadium-dependent bromoperoxidase"
/protein_id="AAM46061.1"
/db_xref="GI:21309840"
/translation="MGI PADNLOSRAKASFDRVSAELALARGVVPVSLANGBELLYR
NPDPNGDSFVSFTKGLPHDDNGAIIDPDDFLAFVRAINSGDEKEIADLTIGPAD
PDTGLPIWRSDLANSLEVRGWENSSAGLTFDLEGPDAQSIAMPAPVLTSPELIAE
IAELYLMALGRIEFSEFSDPSKNAEYIQFAIDQNGLEWTFPAMLPDPPAEIRRRG
EVTNGLFLRIGLPGSEVPGYLSQYIIIVGSKQIGSATGKNTLVSPPNAADFGEIAYG
STISORVRIATPGDRDFTDLKVLVDQAAADPFESYEPGARLRTIRDLATWVHF
DALYEALVACLILLANRPVDPENIPQOEDKLDNDQVFNFGDAHVLSTVETATRA
LKAVRQKNHNRRLRPATGGLISVNTKNAEKGESVFPFVDLAVSELEILEKAEIS
NRKNADGDDPDDPFLPLQPAFAGSPPHPSYSGSHAVVAGACVILKAFPSNFQI
DQVFEYDKEDKLKSSFKGTLTAVAGELANKLADNIAIGRNMGVHVPFSDQFESILLGE
QVAIGILEBSQSLTYGNEFFNLPKPGDTTIQI"
```

## ORIGIN

Alignment Scores: 1.39e-36 Length: 1908  
Pred. No.: 626.00 Matches: 176  
Score: 45.00% Conservative: 76  
Percent Similarity: 31.43% Mismatches: 220  
Best Local Similarity: 17.74% Indels: 88  
Query Match: 8 Gaps: 15  
DB:

US-10-691-383-2 (1-676) x AF218810 (1-1908)

QY 168 AsnGlyGluThrAspLeuAlaThrMetPheHisSerLeuProHisAspGluLeuGly 187  
 |||||  
 Db 148 AACGAGATCAAGCTTCATGCTTACGATGCTTACCAAGGGTCTTCGCGATGACGAAATGGC 207  
 |||||  
 QY 188 Gln---ValThrAlaAspAspPheAlaIleGluAspCysIleLeuAenGlyAspPhe 206  
 |||||  
 Db 208 GCTATTATGACCGGACGACTTTCGGCTTGGTCTGGATCAATAGTGGAGATGAG 267  
 |||||  
 QY 207 SerIleCysGluAspVal-----ProAlaGlyAspProAlaGlyArgLeu----- 221  
 |||||  
 Db 268 AAGGAGATCGCGACCTCACATGGGGCCAGCTCGCGACCGGACACTGGCTTACCAATC 327  
 |||||  
 QY 222 -----ValAsnProThr 225  
 |||||  
 Db 328 TGGCGCTCGGATTTCGCGAATTTCTCGAGCTCGAAGTGGGAGGATGGGAAACAGCTCT 387  
 |||||  
 QY 226 AlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThrThrIleProProVal 245  
 |||||  
 Db 388 GCGGCTCTACCTTCGACCTTGGAGGCGCGAGCGGAGTGGATGCGCATGCCACCGCG 447  
 |||||  
 QY 246 ProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeu 265  
 |||||  
 Db 448 CCTGTGCTCAGGACCTGAGCTCATCGCGGAGATACGAGACTGTACTGATGGCGCTT 507  
 |||||  
 QY 266 AlaArgAspValProPhe-----MetGlnTyr 274  
 |||||  
 Db 508 GGACGCGAATTCGAGTTTAGCGAGTTTTCGATTCGCCAAGAACGACGAGTATATTCAGTTT 567  
 |||||  
 QY 275 GlyThrAspGluIleThrThrAlaAlaAsnLeuAlaGlyMetGlyGlyPhePro 294  
 |||||  
 Db 568 GCTATTGATCAGCTTAAACGGGCTAGAGTGGTTCAACACGCGCGCAATGCTCGGAGATCCG 627  
 |||||  
 QY 295 AsnLeuAspAlaValSerIleGlySerAspGlyThrValAspProPheSerGlnLeuPhe 314  
 |||||  
 Db 628 CTGCGGAATCCGTCGCCCTCGCGTGGAGTGTCTC-----GGAACTTGTTC 678  
 |||||  
 QY 315 ArgAlaThrPheValGlyValGluThrGlyPropheValSerGlnLeuVal----- 332  
 |||||  
 Db 679 CGCGGTATTCTTCAGGCTCTGAGGCTCGGCGCGGTATCTCAGCCAGTACATCATCTGTTGGT 738  
 |||||  
 QY 332 ----- 332  
 Db 739 AGCAAGCAGATTGGCTCCGCGACGGGTGGCAACAAACTCTGTGAGCCCCAATGCGGCT 798  
 |||||  
 QY 333 -----AsnSerPheThrIleAspAlaIleThrValGluProLysGlnGluThr 348  
 |||||  
 Db 799 GATCAGTTTGTATGGTGAATCGCTACGGAAGCATCACCATAGCCAGCGGTGCGTATC 858  
 |||||  
 QY 349 PheAlaProAspLeuAsnTyrMetValAspPheAspGluTrpLeuAsnIleGlnAsnGly 368  
 |||||  
 Db 859 GCCACGCTGGGCGGACCTTATGACCGCACTTGAAGGTATTCCTTGACGTCCAGGACGCT 918  
 |||||  
 QY 369 GlyProAlaGlyProGluGluLeuAspGluLeuArgPheIleArgAsnAlaArg 388  
 |||||  
 Db 919 GCGGACTTCGAGCTTTGATGCTGATGACCGGGAGCAGCGCTCATCCGAGCATCCGC 978  
 |||||  
 QY 389 AspLeuAlaArgValSerPheValAsnIleAsnThrGluAlaTyrArgGlySerLeu 408  
 |||||  
 Db 979 GATCTGCGACGCTGGGTGCACTTTCGACGCTGACGAGCCCTACCTCAATGCTGCTG 1038  
 |||||  
 QY 409 IleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyPropheIleAspSer 428  
 |||||  
 Db 1039 ATTCTGTGGGAACCGCGTGGCTTCGATCCCAACATT-----CGTTTCCAGCAGGAG 1092  
 |||||  
 QY 429 AspArg-----GlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeu 445  
 |||||  
 Db 1093 GACAAGCTCGAATACCGAGGACGTGTTGTGAACCTTCGGAGCCGACACGCTGCTGAGTCTG 1152  
 |||||  
 QY 446 Ile---GlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnIlyTrpGlnValHis 464  
 |||||  
 Db 1153 GTGACTGAGTGGCTACCGCGCGGTGACGCGGTGACCGTACCAAGATTATCATTCAT 1212  
 |||||  
 QY 465 ArgPheAlaArgProGluAlaLeuGlyThrLeuHis---AsnThrIleAlaGlyAsp 483

Db 1213 CGTGCCTCGCGCTGAGGCTACCGGTGGTCTGATTAGCTGAACAAATTCGACGCGAG 1272  
 |||||  
 QY 484 -----LeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeu 499  
 |||||  
 Db 1273 AAGGCGGAGAGCGTTTCCCTGAGTGTGACTGCTGCTTGAAGAGCTTGAAGATATCTGT 1332  
 |||||  
 QY 500 LysArgValAlaGluIleAsnAlaAlaGlnAsn-----ProAsnAsnGlu 514  
 |||||  
 Db 1333 GAGAAAGCTGAATTAGCAATAGGAAACAGAACATTGCTGACGAGAGATCTCTGACCTGAT 1392  
 |||||  
 QY 515 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 534  
 |||||  
 Db 1393 CTTTCATTCTGTTGCGCAGGCAATTCGCCAGGCGGAGCCATTCCTCCGCTCTACGGA 1452  
 |||||  
 QY 535 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGly--- 553  
 |||||  
 Db 1453 AGCGGCCACGCTGTTGCTGGCGCATGTGTGACGATCTCAAGGGGTTCTTCGACTCC 1512  
 |||||  
 QY 554 -----LeuAspArgGlyGlyGluCysPheProAsnProValPheProSerAspAsp 570  
 |||||  
 Db 1513 AACTTCCAGATCGATCAGGTGTTTCGAG-----GTCGACAAAGATGAGGAC 1557  
 |||||  
 QY 571 GlyLeuLeuLeuLeuAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnLys 590  
 |||||  
 Db 1558 AAGCTTGTAAAGTCGCTTTTCAAGGGAACCT---CTCACTGTTGCCGCTGAGTGAACAAG 1614  
 |||||  
 QY 591 LeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGly 610  
 |||||  
 Db 1615 CTGCGCGACATATTCGATCGCGGGAACATGCGAGTGTCTCACTCTCTCTGACCCAG 1674  
 |||||  
 QY 611 IleGlnGlyLeuLeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMet 630  
 |||||  
 Db 1675 TTCGAGTCAATTCCTGCTTGTGAGCAGTTTCGATCGAATCTTGAAGAGACGAGTCTG 1734  
 |||||  
 QY 631 ThrPheAlaGluAlaThrPheGluPheArgLeuPheThrGlyGluValIleLysLeu 650  
 |||||  
 Db 1735 ACGTATGGCGAGAATCTTCTTCAACTTCGCGAAGTTTGTGTAACATACTCAATCCAGATC 1794  
 |||||  
 RESULT 9  
 LOCUS D87657 2035 bp mRNA linear PLN 07-JUL-1998  
 DEFINITION Corallina pilulifera mRNA for vanadium-dependent bromoperoxidase 1,  
 complete cds.  
 ACCESSION D87657  
 VERSION D87657.1 GI:3293049  
 KEYWORDS vanadium-dependent bromoperoxidase 1.  
 SOURCE Corallina pilulifera  
 ORGANISM Corallina pilulifera  
 Eukaryota; Rhodophyta; Florideophyceae; Corallinales;  
 Corallinaceae; Corallinoideae; Corallina.  
 REFERENCE 1 (bases 1 to 2035)  
 AUTHORS Shimomishi,M., Kuwamoto,S., Inoue,H., Wever,R., Ohshiro,T.,  
 Izumi,Y. and Tanabe,T.  
 TITLE Cloning and expression of the gene for a vanadium-dependent  
 bromoperoxidase from a marine macro-alga, Corallina pilulifera  
 JOURNAL FEBS Lett. 428 (1-2), 105-110 (1998)  
 MEDLINE 98307393  
 PUBMED 9645486  
 REFERENCE 2 (bases 1 to 2035)  
 AUTHORS Tanabe,T.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-SEP-1996) Tadashi Tanabe, National Cardiovascular  
 Center Research Institute, Department of Pharmacology;  
 Fujishiro-dai 5-7-1, Suita, Osaka 565, Japan  
 (E-mail:tanabeori.ncvc.go.jp, tel:06-833-5012, Fax:06-872-8092)  
 FEATURES  
 Location/Qualifiers  
 1..2035  
 /organism="Corallina pilulifera"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:78447"  
 /clone="BP01"  
 128..1924  
 CDS

```

/codon_start=1
/product="vanadium-dependent bromoperoxidase 1"
/protein_id="BA331261.1"
/db_xref="gi:3293050"
/translation="MGIPADNLSQRAKASFDTRVAALALNRGVVPSFANGBELLYR
NPDDNTDPSFIASFYTKGLPHDNGAIIDPDDFLAFVRAINSGDKEIADLTIGPARD
PETGLPTMRDLANSLELEVRGWENSAGLTFDLEQDAOSIAMPPAPVLTSPELVAE
IAELYLMALGRETEFEFSDSPKNAEVIQPAIDOLNGLEWENTPKLGDPPAEIRRRRG
EVTGNLFRGILPGSVGPYLSQYIIVGSKQIGSATVGNKTLVSPNADEFDEIAYG
SITISQVRITPERGFMIDKLVLDVQDAADFPGFSEYEPGARLIRTLKDLAIWHF
DALYEAFLNACLILANGVEFDNPLFPQDDKLDNQDFVFNFSAHVLSLTVAVTRA
LKAVYQKNIENRPGATGGLISVNIKAPQGESIFPEVDLAVBELGDILEKARIS
NRKONTADGDDPDSFLLPMAFAGSPHPSPSGHVVAGACVILKAFPSGIEI
DOVFEVDKDDKLVKSSFKGLTVAGELNKLADNIALGRNMGVHVFSDQFESLLIGE
QVAIGLIEQSLTYGENFPNLPKPDGTIQL"

```

## ORIGIN

## Alignment Scores:

```

Pred. No.: 1.52e-36 Length: 2035
Score: 626.00 Matches: 194
Percent Similarity: 43.77% Conservative: 73
Best Local Similarity: 31.80% Mismatches: 239
Query Match: 17.74% Indels: 104
DB: 8 Gaps: 20

```

US-10-691-383-2 (1-676) x D87657 (1-2035)

```

QY 135 ProThrAlaProAsnArgArgAspAsnVal-----AlaPheAlaSerArg 149
DB 122 CCTACATGGGTATTCAGCTGACACCTCCAAAGTCGCGCCCAAGGCTTCATTCGATACG 181
QY 150 ArgAspAlaAlaArgArgGluArgAspGlyThrGlyThrValCysGlnIleThrAsnGly 169
DB 182 CGTGTAGCTCGGCGAGCTTGACCTCAACCGCGAGTTGCATCGCTTTGCAAAATGGG 241
QY 170 Gln-----ThrAsp-----LeuAlaThrMet 176
DB 242 GAAGAGCTTCTACCGCAACCGGACCTCGACACACTGATCCGAGCTTCATCGCTAGC 301
QY 177 PheHisLysSerLeuProHisAspGluLeuGlyGln---ValThrAlaAspAspPheAla 195
DB 302 TTCACAAAGGCTCTTCGCGATACGCAATAGGCGCTATTATCGACCCCGAGCACTTCTTG 361
QY 196 IleLeuGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAspVal----- 212
DB 362 GCCTTCGTCGTCATCAATAGTGGCGATGAAGAGAGATCGCCGACCTCACATTCGGG 421
QY 213 ProAlaGlyAspProAlaGlyArgLeu----- 221
DB 422 CCAGCTCGGACCGGAGAGCTGGCTTACCAATCTGGCGCTCGGATCTGGCGAAATTCCTC 481
QY 222 -----ValAsnProThrAlaAlaPheAlaIleAspIleSerGly 234
DB 482 GAGCTCGAAGTGGCAGATGGGAGACAGCTCTGCGCGTCTCACCTTCGACCTGGAGGGC 541
QY 235 ProAlaPheSerAlaThrIleProProValProThrLeuSerSerProGluLeuAla 254
DB 542 CCGAGCGGAGTGGATTGCCATGCCACCGCGCTGTGTCTACGAGCCCTGAGCTCGTC 601
QY 255 AlaGlnLeuAlaGluLeuIleThrMetAlaLeuAlaArgAspValProPhe----- 271
DB 602 GCCGAGATAGCAGAGCTGTACTGTATGGCGCTTGGACGCGAAATCGAGTTTCAGCGAGTTT 661
QY 272 -----MetGlnTyrGlyThrAspGluIleThrThrAla 283
DB 662 GATTCCCCAAAGAACCGCAGAGTATATTCAGTTTGCTATTGATCATGCTTAACGGGCTGGAG 721
QY 284 AlaAlaAsnLeuAlaGlyMetGlyPheProAsnLeuAspAlaValSerIleGlySer 303
DB 722 TGGTTCAACACACCGGCAAGCTCGGAGATCCGCTCGGAAATCCGTCGCGCGCGGT 781
QY 304 AspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGlyValGluThr 323
DB ::::|

```

```

DB 782 GAGGTGACTGTT-----GGAAACTTGTTCGCGGTATTCTTCCAGGCTCTGAGGTC 832
QY 324 GlyProPheValSerGlnLeuLeu-----ValAsnSerPheThrIle 337
DB 833 GGCCCGTACTCGACGAGTACATCATCGTTGGTAGCAAGCAGATGGCTCAGCAGAGTT 892
QY 337 ----- 337
DB 893 GGTAAACAAACTCTCGTGAGCCCAATGCTGCTGATGAGTTTGTGTTGATGAAATCGCCTAC 952
QY 338 AspAlaIleThrValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrMetVal 357
DB 953 GGAAGCATCACCATTACCCAGCGTGTGCGTATCCGCCAGCGTCGACGCGCATTCATGACC 1012
QY 358 AspPheAspGluThrLeuAsnIleGlnAsnGlyGlyProAlaGlyProGluGluLeu 377
DB 1013 GACTTGAAGGTATTCCTTGACGTCCAGGACGCTCGGACCTTCGAGGCTTTGAGTCGTAT 1072
QY 378 AspGluGluLeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPheValAsp 397
DB 1073 GAGCCGGGAGCAGCGCTCATCCGAGCATCCGCACTTCGCGAGTGGTGCACTTTGAC 1132
QY 398 AsnIleAsnThrGluAlaTyrArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSer 417
DB 1133 GCACTGTACGAGGCTACCTCAATGCGTGCCTAATCTGTTGGCAACGCGCTGCCGTTTC 1192
QY 418 ArgProGlyIleAsnGlyProPheIleAspSerAspArg-----GlnAlaGlyPhe 434
DB 1193 GATCCC-----AACCTTCGTTCCAGGAGGAGCAAGCTCGATACACGAGGAGTGT 1246
QY 435 ValAsnPheGlyThrSerHisTyrPheArgLeuIle---GlyAlaAlaGluLeuAlaGln 453
DB 1247 GTCAACTTCGATCCGACACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1306
QY 454 ArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeuGly 473
DB 1307 AAGCGGTACGTTACCAAGAGTTTAAACATTCATCGTCCGCTCGGCGCTGAGGCTACCGGT 1366
QY 474 GlyThrLeuHis---AsnThrIleAla-----GlyAspLeuAspAlaAspPhe 488
DB 1367 GGTCTGATTAGCGTTTACAAAATCCGACCCGAGAGGCGAGAGCATTTTCCCTGAGGTT 1426
QY 489 AspIleSerLeuLeuAsnAspGluLeuLeuLysArgValAlaGluIleAsnAlaAla 508
DB 1427 GATCTTCTGCTTCAAGAGCTTGAGATATCTTCGAGAAAGCTGAAATAGCAATAGGAAG 1486
QY 509 GlnAsn-----ProAsnAsnGluValThrTyrLeuLeuProGlnAlaIle 523
DB 1487 CAGAACATAGCTGACGAGATCCTGACCTGATCCTTCATTCCTGTTGCGCATGGCAATC 1546
QY 524 GlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAla 543
DB 1547 GCGAGGCGGACCCCATTCATCCGTCCTACGGAAGCGGCCACGCTGTGTTGTTGGCGCA 1606
QY 544 PheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyCysPheProAsn 563
DB 1607 TGTGTGACGATCTGAAGCGGTTTC-----TTCGACTCCGCGCATCGAGATC-----GAT 1654
QY 564 ProValPheProSerAsp-----AspGlyLeuGluLeuLeuAsnPheGluGlyAla 580
DB 1655 CAGGTGTTCGAGTTCGACAAAGATGAGGCAAGCTTGTGAAGTCGCTTTTCAAGGAACT 1714
QY 581 CysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGln 600
DB 1715 ---CTCATTCTTCGCGTGAATTAACACAGCTCGCCGACAAATATGCGATCGGCGGTAAAC 1771
QY 601 MetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuLeuGlyGluThrIle 620
DB 1772 ATGCGAGGTGTTCACTACTTCTCTGACCACTTCGAGTCACTTCTGCTCGGTGAGCAGGTT 1831
QY 621 ThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluAlaThrPheGluPhe 640
DB 1832 GCGATTGGAATCTTGAAGAGCAAGTCTGACGTATGCGGCAAGACTTCTTCTTCAACTGT 1891

```

```

Qy 641 ArgLeuPheThrGlyGluValIleLeu 650
Db 1892 CCAGAGTTTGATCGAAGTACAATCCAGATC 1921

RESULT 10
LOCUS E17199 1794 bp DNA linear PAT 28-JUL-1999
DEFINITION Corallina pilulifera mRNA for haloperoxidase.
ACCESSION E17199
VERSION E17199.1 GI:5711882
KEYWORDS JP 1998248581-A/1.
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 1794)
AUTHORS Izumi, Y. and Tanabe, T.
TITLE NEW HALOPEROXIDASE GENE AND ITS UTILIZATION
JOURNAL Patent: JP 1998248581-A 1 22-SEP-1998;
COMMENT OTSUKA PHARMACEUT CO LTD
PN JP 1998248581-A/1
PD 22-SEP-1998
PF 06-MAR-1997 JP 1987070539
PI IZUMI YOSHIKAZU, TANABE TADASHI
PC C12N15/09, C12N9/08, (C12N15/09, C12R1:89), (C12N9/08, C12R1:19);
CC strandedness: Double;
CC topology: Linear;
FH Key Location/Qualifiers
FH source 1..1794
FH mat_peptide /organism='Corallina pilulifera' FT
FT /clone='BP01'
FT /product='holoperoxidase'.
FEATURES
source Location/Qualifiers
1..1794
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

ORIGIN
Alignment Scores:
Pred. No.: 2,12e-36 Length: 1794
Score: 623.00 Matches: 191
Percent Similarity: 44.41% Conservative: 75
Best Local Similarity: 31.89% Mismatches: 233
Query Match: 17.66% Indels: 100
DB: 6 Gaps: 20

US-10-691-383-2 (1-676) x E17199 (1-1794)
Qy 141 ArgAspAsnValAlaPheAlaSerArgArgAspAlaAlaArgArgGluArgAspGlyThr 160
Db 31 CGGCCAAGGCTTCATTCGATACCGCGGTAGCTCGCGCGAGCTTGCACTCAAC---CGC 87
Qy 161 GlyThrValCysGlnIleThrAsnGlyGlu 170
Db 88 GGAGTTGTACCATCGTTTGCAATGGGAGAGAGCTTCTTACCCCAACCCGGACCTGAC 147
Qy 171 ---ThrAsp-----LeuAlaThrMetPheHisLysSerLeuProHisAspGluLeuGly 187
Db 148 AACACTGATCCGAGCTTCATCGCTAGCTTCACAAGGGTCTTCGCGCATGACGACAATGGC 207
Qy 188 Gln---ValThrAlaAspAspPheAlaIleLeuGluAspCysIleLeuAsnGlyAspPhe 206
Db 208 GCTATTATCGACCCGACGACTTCTTGGCTTCGTCGCAATCAATAGTGGCGATGAA 267
Qy 207 SerIleCysGluAspVal-----ProAlaGlyAspProAlaGlyArgLeu----- 221
Db 268 AAGGAGATCCCGACCTTCACATGGGGCCAGCTCGCGACCGGAGACTGCGTTACCAATC 327
Qy 222 -----ValAsnProThr 225

```

```

Db 328 TGGCGCTCGGATCTGGCGAAATTTCTCTCGAGCTCGAAGTGCAGGATGGGAGACAGACTCT 387
Qy 226 AlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThrThrIleProProVal 245
Db 388 GCCGCTCTACCTTCGACCTGGAGGCGCGGACCGGAGTCGATGCCATGCCACCGCG 447
Qy 246 ProThrLeuSerSerProGluLeuAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeu 265
Db 448 CCGTGTCTCAGAGCCCTGAGCTCGTCCGCGAGATAGCAGAGCTGTACTCATGGCGCTT 507
Qy 266 AlaArgAspValProPhe-----MetGlnTyr 274
Db 508 GGACGCGAAATCGAGTTTCAGCGAGTTTTCATCCCAAGAACGCGAGAGTATATTCAAGTT 567
Qy 275 GlyThrAspGluIleThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPhePro 294
Db 568 GCTATTGATCAGCTTAACGGGCTGAGTGGTTCAACACCGCGGAAAGCTCGGAGATCCG 627
Qy 295 AsnLeuAspAlaValSerIleGlySerAspGlyThrValAspProPheSerGlnLeuPhe 314
Db 628 CCGCGGAAATCCGTCGCCGCGGTGAGGTGACTGTT-----GGAAACTTGTTC 678
Qy 315 ArgAlaThrPheValGlyValGluThrGlyProPheValSerGlnLeuLeu----- 331
Db 679 CGCGGTATTCTTCAGGCTCTGAGTGCGCCGCTACCTCAGCCAGTACATCATCGTTGGT 738
Qy 332 -----ValAsnSerPheThrIle----- 337
Db 739 AGCAGCAGATTTGGCTCAGCGACAGTTGGTTAAACAAACTCTCGTGAGCCCCAATGCTGCT 798
Qy 338 -----AspAlaIleThrValGluProLysGlnGluThr 348
Db 799 GATGAGTTTGATGGTGAATCGCCTACGGAAGCATCACCATTAGCCAGCGTGCATC 858
Qy 349 PheAlaProAspLeuAsnTyrMetValAspPheAspGluTrpLeuAsnIleGlnAsnGly 368
Db 859 GCCACGCTCGGACGCGACTTCATGACGCGACTTGAAGGTATTCTCTGACGTCGAGCGCT 918
Qy 369 GlyProProAlaGlyProGluGluLeuAspGluLeuArgPheIleArgAsnAlaArg 388
Db 919 CGGAGCTTCGAGCTTTGAGTCGTATGCGCGGAGCAGCGCTCATCCGACCATCGC 978
Qy 389 AspLeuAlaArgValSerPheValAspAsnIleAsnThrGluAlaIleArgGlySerLeu 408
Db 979 GATCTCGCAGCTGGGTGCCTTTGACGCACCTGACGAGGCTACCTCAATGCGTGCCTA 1038
Qy 409 IleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSer 428
Db 1039 ATTCTGTTGGCAACGCGCTGCGCTTCGATCCC-----AACCTTCGCTTCAGCAGGAG 1092
Qy 429 AspArg-----GlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeu 445
Db 1093 GACAACTCGATACCAAGGACGCTGTTGTGAACCTCCGATCCGACAGCTGCTGAGTCTG 1152
Qy 446 Ile---GlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnValHis 464
Db 1153 GTGACTGAAGTGCTACGCGCGGTTGAAGCGGTACGCTACGAGAGATTAAACATTCAT 1212
Qy 465 ArgPheAlaArgProGluAlaLeuGlyGly---ThrLeuHisAsnThrIleAla----- 481
Db 1213 CGTCGCTGCGCTCGAGGCTACGCGTGGTCTGATTATCGTTTAAACAAATCGCACCGCAG 1272
Qy 482 -----GlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeu 499
Db 1273 AAGGCGAGAGCATTTTCCTCGAGTTGATCTTCTGCTGTGAAGAGCTTGGAGATATCTTG 1332
Qy 500 LysArgValAlaGluIleAsnAlaAlaGlnAsn-----ProAsnAsnGlu 514
Db 1333 GAGAAAGCTGAATTAGCATAGGACGACAGACATAGCTGACGAGAGATCTCTGACCTGAT 1392
Qy 515 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 534

```

Db 1393 CTTTCATTCCTGTTGCGGATGCGCATTCGCGAGGAGCCCATTCATCCGTCCTACGGA 1452

Qy 535 SerGlyHisAlaThrGlnAnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeu 554  
|||||  
|||||

Db 1453 AGCGGCCACCTGCTGGTGTGGCGCATGTGTGACGATCCTGAAGCGGTTTC-----TTC 1506

Qy 555 AspArgGlyGlyGluCysPheProAsnProValPheProSerAsp-----AspGly 571  
|||  
|||

Db 1507 GATCCGGCATCAGATC-----GATCAGGTGTTGAGGTGCACAAAGATGAGGACAG 1560

Qy 572 LeuGluLeuLeuLeuAsnGluGlyAlaCysLeuThrTyrGluGlyGluLeuLeuLeu 591  
|||  
|||

Db 1561 CTTGTGAAGTCGCTTTTCAAGGAACT---CTCACTGTTCCCGGTGAATTAAGCAAGCTC 1617

Qy 592 AlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIle 611  
|||  
|||

Db 1618 GCGCAACAATATTCGATCGCGCGTAACATGCGAGGTGTTCACTACTTCTCTGACCAAGTTC 1677

Qy 612 GlnGlyLeuLeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThr 631  
|||  
|||

Db 1678 GATCACTTCTGCTCGGTGAGCAGGTGCGATTTGGATCTTTGGAAGAGCAAGTCTGACG 1737

Qy 632 PheAlaGluGluAlaThrPheGluPheArgLeuPheThrGlyGluValIleLysLeu 650  
|||  
|||

Db 1738 TATGGCGAGAACTTCTTCTCAACTTGCAGAGTTTGATGGAACACTACAATCCAGATC 1794

RESULT 11

Db 87658

LOCUS

DEFINITION Corallina pilulifera mRNA for vanadium-dependent bromoperoxidase 2, complete cds.

ACCESSION DB87658

VERSION DB87658.1 GI:3293051

KEYWORDS vanadium-dependent bromoperoxidase 2.

SOURCE Corallina pilulifera

ORGANISM Corallina pilulifera  
Eukaryota; Rhodophyta; Florideophyceae; Corallinales;  
Corallinaceae; Corallinoideae; Corallina.

REFERENCE 1 (bases 1 to 2029)  
Shimomishi, M., Kuwamoto, S., Inoue, H., Wever, R., Ohshiro, T., Izumi, Y. and Tanabe, T.  
Cloning and expression of the gene for a vanadium-dependent bromoperoxidase from a marine macro-alga, Corallina pilulifera  
FEBS Lett. 428 (1-2), 105-110 (1998)

JOURNAL 98307393

MEDLINE 9645486

REFERENCE 2 (bases 1 to 2029)  
Tanabe, T.  
Direct Submission  
Submitted (02-SEP-1996) Tadaishi Tanabe, National Cardiovascular Center Research Institute, Department of Pharmacology,  
Fujishiro-dai 5-7-1, Suita, Osaka 565, Japan  
(E-mail: tanabe@ri.ncvc.go.jp, Tel: 06-813-5012, Fax: 06-872-8092)  
Location/Qualifiers

FEATURES

source

1. 2029  
/organism="Corallina pilulifera"  
/mol\_type="mRNA"  
/db\_xref="taxon:78447"  
/clone="BPO2"  
119..1912  
/codon\_start=1  
/product="vanadium-dependent bromoperoxidase 2"  
/protein\_id="BA331262.1"  
/db\_xref="GI:3293052"  
/translations="MGIPADNLOSRAKASFDTRVAAAEALALARGVVPSPFANGBELYR  
NCTGSPSTASTTGLHDDNGAIIDPDDFLAFVRAINSQDEKEADLTGLPARDPE  
TGLPIWRSLDLELEVRGWENSSAGLTFDLEGPQAQSVAMPAPVLPMSPELIASMA  
ELYLMAGRDIEFSESPKNAAFIRSAIERLNGLEWNTPAKLGPPPARIRRRGEV  
TGNLFRGLIPSEVGPYLSQYILVGSQKIGSATVGNKTVSPNADEPGEIAYGSI  
TISQVRATPGDFTMDLFDVLDVQDQDFRGFESFEPGARLIRTIIRDALTWHFADK  
LISAYLNACILILANGVFPDLPFOQEDKLDHNVFNFGSAHVLISVTEVATRAIK  
AVRYQKFNIRHRLRPLAGGLISVNNKSLAGSDIIFPEVSELVSLSDDDVAESN  
EKQNRADGIVSPDKSFLPMAFAEGSPFHPSPSGSHAVAGACVTILKAPFDANFQID

CDS

## ORIGIN

Alignment Scores: 6.35e-36 Length: 2029  
Pred. No.: 617.50 Matches: 201  
Score: 41.12% Conservative: 77  
Percent Similarity: 41.12% Mismatches: 229  
Best Local Similarity: 29.73% Indels: 169  
Query Match: 17.50% Gaps: 24  
DB: 8

US-10-691-383-2 (1-676) x D87658 (1-2029)

Qy 67 ProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIleMetAlaArg-----Val 84  
|||||  
|||||

Db 113 CCACAAATGGGTATTCAGCTGCAACCTCAAAGTCGCGCCCAAGGCTTCATTCGATACG 172

Qy 85 ArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPheAspGluVal 104  
|||  
|||

Db 173 CGTGTAGCTCGGCC-----GAGCTGGCA 196

Qy 105 ThrAlaSerGlyValPheProGluGluHisLysHisThrGlyGluGlyArgHisLeuGln 124  
|||  
|||

Db 197 CTCGCCGTGGAGTTGTGCCA----- 217

Qy 125 ThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArgArgAspAsnVal 144

Db 217 ----- 217

Qy 145 AlaPheAlaSerArgArgAspAlaAlaArgGluArgAspGlyThrGlyThrValCys 164  
|||  
|||

Db 218 TCGTTTGCATAATCGAGGAGCTTCTTACCGCAAC-----TGT 256

Qy 165 GlnIleThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSerLeuProHisAsp 184  
|||  
|||

Db 257 GAG-----ACCGCGCATCCAAGCTTCAATGCTAGCTTCAAAAGGCTTCCACATGAC 310

Qy 185 GluLeuGlyGln---ValThrAlaAspPheAlaIleLeuGluAspCysIleLeuAsn 203  
|||  
|||

Db 311 GACAATGGCGCTATTATTATTCGCGAGATTTCTTGGCTTCTGTCGTGGATCAATAGT 370

Qy 204 GlyAspPheSerIleCysGluAspVal-----ProAlaGlyAspProAlaGlyArg 220  
|||  
|||

Db 371 GGAGATGAGAGGAGATCGCGGACCTTACATTCGGGCGCAGCTCGCGACCTCGAGATGGC 430

Qy 221 Leu-----Val 222

Db 431 TTACCAATCTGGCGCTCAGATCTGGCGAATTTCTCGAGCTCGAAGTCGAGGATGGAA 490

Qy 223 AsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThrIle 242  
|||  
|||

Db 491 AACAGCTCTCGCGGTCTTACCTTCGACCTGGAGGCGCGGACGCGGAGTGGTGGCCATG 550

Qy 243 ProProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrp 262  
|||  
|||

Db 551 CCACCGCGCTGTGCTCATGAGCCCTGAGCTCATCGCGAGATGGCAGAACTGATCCTG 610

Qy 263 MetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGlu-----Ile 279  
|||  
|||

Db 611 ATGGCGCTTGGAGCGCACATCGAGTTTAGCGAGTTTCAGAGTCCCAAGAACGACGCTTTT 670

Qy 280 ThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAla--- 298  
|||  
|||

Db 671 ATTAGGTCTGCTATTGAACGGCTTAACGGGCTGAGTGGTTC---AACACACCGGCAAG 727

Qy 299 -----ValSerIleGlySerAsp 304  
|||  
|||

Db 728 CTTGGTATCCGCTCGGGAATCCCGTCGCGGTGAGGTGAGTGCAGCGTGGAAAC--- 784

Qy 305 GlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGlyValGluThrGly 324  
|||  
|||

Db 785 -----TTGTTCCGCGGTATTCTTCCCGGTTCCTCCCGGTTCGAGGTCGCG 820



```
Qy 325 ProPheValSerGlnLeuLeu-----ValAsnSerPheThrIle--- 337
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 821 CCGTATCTCAGCCAGTACATCATCGTTGGTAGCAAGCAGATTGGCTCTCGCAGCGTTGGC 880
338 -----
Qy 338 -----Asp 338
881 AACAAAACTTTCGTGAGCCCCAATGCGCGTGATGAATTTGATGTGAAATCGCCTACGGA 940
Qy 339 AlaIleThrValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrMetValAsp 358
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 941 AGCATCACCATTACCCAGCGTGTGGTATCGCCACGCGCTGGCGCGACTTCATGACTGAT 1000
359 PheAspGluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAsp 378
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1001 TTGAAGGTATTCCTTGAGTCCAGTCCAGACGGTGCAGACTTCGAGGATTTGAGTCGTATGAG 1060
379 GluGluLeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsn 398
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1061 CCGGAGCAGCGCCATCTCGGACGATCCGCGATCTTCGACGCTGGGTGCACATTTGACGCA 1120
399 IleAsnThrGluAlaTyrArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSerArg 418
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1121 CTGTACAGGCTACCTCAATCGTGTGCTGATTCCTGTTGGCAACGCGGTCCGTTTCGAT 1180
419 ProGlyIleAsnGlyProPheIleAspSerAspArgGlnAlaGly-----PheVal 435
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1181 CCC-----AACCTTCCTTCCAGCAGGAGGACAACTCGACCAACCAGATGATGTTGTG 1234
436 AsnPheGlyThrSerHisTyrPheArgLeuIle---GlyAlaAlaGluLeuAlaGlnArg 454
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1235 AACTTCGGATCCGACACAGCTGCTGAGTCTGTGATCTGAAAGTGGCTGCGCGCGTTGAAG 1294
455 AlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGly 474
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1295 CGGTACGGTATCAGAAGTTTAACTTCATTCCTCGCTCGCGCTGAGGCTTACTGGTGGT 1354
475 ThrLeu-----HisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIle 490
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1355 CTGATTAGCTTAACAAGAAATCTTCTTCGCGGTAGT-----GATATA 1399
491 SerLeuLeuGluAsnAspGluLeuLeuLysArg-----ValAla 503
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1400 ATATTCCTCGAAGTTAGTAGTACTGCTGCGAGAGCTCTCATCAATTCGTGATGACGTTGCT 1459
504 GluIleAsnAlaAlaGlnAsnProAsnAsnGluVal-----ThrTyrLeu 518
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1460 GAGAGCAATGMAAACAGAACAGGCGCTGACGGGATTTGTAGCCCGCGATTAATCAATTCGT 1519
519 LeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAla 538
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1520 TTGCGGATGGCATTTGCGAGGCGAGCCCATTCCTCGTCTATGGAAGTGGCCACGCT 1579
539 ThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGly-----Leu 554
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1580 GTGTTGCTGGCGCATGTGTGACAATCTCAAGCGCTTCTTCGACGCGCAACTTCAGATC 1639
555 AspArgGlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeu 574
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1640 GATAAGTGTTCGAG-----GTCGACACTGATGAGGACAAGCTTGTGAAG 1684
575 IleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsn 594
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1685 TCGCTTTTCAAGGGAAGCT---CTCACTGTTGCCGTGAATTGAACAAGCTCGCCGACAT 1741
595 ValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeu 614
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1742 GTTCGATCGGGCGGAACATGGCGGTGTTCACTACTTCTCTGACCAAGTTTCGATCACTG 1801
615 LeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGlu 634
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1802 CTACTCGGTGAACAGATTGGGAAATCTTGGAGGAGCAGACCTTGACGTATGGCGAG 1861
```

```
Qy 635 GluAlaThrPheGluPheArgLeuPheThrGlyGluValIleLysLeu 650
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1862 AACTTCTTCTCAATTTGCGAAATTCGACGGAACACTACAATCCAGATT 1909

RESULT 12
E17200
LOCUS E17200 1791 bp. DNA linear PAT 28-JUL-1999
DEFINITION Corallina pilulifera mRNA for haloperoxidase.
ACCESSION E17200
VERSION E17200.1 GI:5711883
KEYWORDS JP 1998248581-A/2.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 1791)
AUTHORS Izumi, Y. and Tanabe, T.
TITLE NEW HALOPEROXIDASE GENE AND ITS UTILIZATION
JOURNAL Patent: JP 1998248581-A 2 22-SEP-1998;
COMMENT OTSUKA PHARMACEUT CO LTD
OS Corallina pilulifera
PN JP 1998248581-A/2
PD 22-SEP-1998
PF 06-MAR-1997 JP 1997070539
PI IZUMI YOSHIKAZU, TANABE TADASHI
PC C12N15/09,C12N9/08,(C12N15/09,C12R1:89),(C12N9/08,C12R1:19);
CC strandedness: Double;
CC topology: Linear;
FH Key Location/Qualifiers
FT source 1..1791 /organism='Corallina pilulifera' FT
FT /clone='BPO2' /mat_peptide 1..1791
FT mat_peptide /product='haloperoxidase'.
FEATURES
source 1..1791
Location/Qualifiers
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

ORIGIN
Alignment Scores:
Pred. No.: 8.9e-36 Length: 1791
Score: 614.50 Matches: 193
Percent Similarity: 43.04% Conservative: 70
Best Local Similarity: 31.59% Mismatches: 223
Query Match: 17.42% Indels: 125
DB: 6 Gaps: 21

US-10-691-383-2 (1-676) x E17200 (1-1791)
Qy 141 ArgAspAsnValAlaPheAlaSerArgArgAspAlaAlaArgGluArgAspGlyThr 160
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 31 CGCGCCAAAGCTTCATTCGATACGCGTGTAGTCGCGCGCGGACTC---GCCCGT 87
161 GlyThrValCysGlnIleThrAsnGlyGlu----- 170
88 GGAGTTGTGCCATCGTTTGCATGAGAGAGAGCTTCTTACCGCAACTGTGACACCGGC 147
171 ---ThrAspLeuAlaThrMetPheHisLysSerLeuProHisAspGluLeuGlyGln--- 188
148 GATCCAAAGCTTCATTCGATGCTTCCACAAAGGCTTCCACATGACCAATGGCGCTATT 207
189 ValThrAlaAspPheAlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIle 208
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 208 ATTGATCCCGACGATTTCCTTGGCGCTTCGTCGCGATCAATAGTGAGATGAGAAGAG 267
209 CysGluAspVal-----ProAlaGlyAspProAlaGlyArgLeu----- 221
268 ATCGCCGACCTTACATTGGGGCCAGCTCGGACCTCGAGACTGGCTTACCAATCTGGCGC 327
222 -----ValAsnProThrAlaAla 227
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Db 328 TCAGATCGCGAATTCTCTGAGCTCGAAGTGGAGATGGGAAAAACAGCTCTGCCGGT 387  
 Qy 228 PheAlaIleAspIleSerGlyProAlaPheSerAlaThrThrIleProValProThr 247  
 Db 388 CTTACCTTCGACCTGGAGGCGCGGAGCGGAGTGGTTCGATCCACCGCGCCCTGTG 447  
 Qy 248 LeuSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAlaArg 267  
 Db 448 CTCATGAGCCCTGAGCTCATCGCCGAGATGGCAGAACTGTACCTGATGGCGCTGGAGCG 507  
 Qy 268 AspValProPheMetGlnTyrGlyThrAspGlu-----IleThrThrThrAlaAla 284  
 Db 508 GACATCAGATTAGCAGATTTCAGTCCCAAGAACGACGCTTTTATTAGTCTGCTATT 567  
 Qy 285 AlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAla----- 298  
 Db 568 GAACGGCTTAACGGCTGGAGTGGTTC---AACACACCGCGCAAGCTTGTGTATCCGCT 624  
 Qy 299 -----ValSerIleGlySerAspGlyThrValAspPro 309  
 Db 625 GCGGAATTCGTCGCCGTCGGGTGAGTGCACGTCGGAAC----- 666  
 Qy 310 PheSerGlnLeuPheArgAlaThrPheValGlyValGluThrGlyProPheValSerGln 329  
 Db 667 -----TTGTTCCGCGGTATTCTTCCCGGTCTCGAGTGGCGCGGTATCTCAGCCAG 717  
 Qy 330 LeuLeu-----ValAsnSerPheThrIle----- 337  
 Db 718 TACATCATGTTGTAGCAAGCAGATTGGCTCTCGACGCTGGCAACAACTTTCGTG 777  
 Qy 338 -----AspAlaIleThrValGlu 343  
 Db 778 AGCCCCAATGTCGCGATGAATTGATGTTGATGTCGAATCGCTACCGAAGCATCACCAATTAGC 837  
 Qy 344 ProLysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPheAspGluTrpLeu 363  
 Db 838 CAGCGTGGGTATCGCCACCGCTCGGCGGACTTCATGACTGATTTGAAGGTATTCCTT 897  
 Qy 364 AsnIleGlnAsnGlyGlyProProAlaGlyProGluLeuAspGluGluLeuArgPhe 383  
 Db 898 GAGCTCAACAGCGGTGAGACTTCCGAGGATTTGAGTCGTATGAGCGGAGCAGCGCTC 957  
 Qy 384 IleArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsnThrGluAla 403  
 Db 958 ATTCCGACGATCCCGATCTTCGACGTGGGTGCATTTGACGCACTGTACGAGCCCTAC 1017  
 Qy 404 TyrArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGly 423  
 Db 1018 CTCATTCGTCCTGATCTCTGTTGGCGAAGCGGCTGCTGATCC-----AACCTT 1071  
 Qy 424 ProPheIleAspSerAspArgGlnAlaGly-----PheValAsnPheGlyThrSer 440  
 Db 1072 CCTTTCAGCAGGAGGACAACTCGAACCATGATGTGTTGTGAACCTTCGATCCGCA 1131  
 Qy 441 HisTyrPheArgLeuIle---GlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGln 459  
 Db 1132 CACGTGCTGAGTCTGGTGACTGAAGTCTACGCGCGCTTGAAGCGGTACGGTATCAG 1191  
 Qy 460 LysTrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThrLeu----- 476  
 Db 1192 AAGTTTAACTTATCTCGCTCGCGCTCGAGCTACTGCTGCTGATTCGTTAATC 1251  
 Qy 477 ---HisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsn 495  
 Db 1252 AAGAAATCTTCTTCCTGGGGTAGT-----GATATAATATTCCTCCCTGAAGTT 1296  
 Qy 496 AspGluLeuLeuLysArg-----ValAlaGluIleAsnAlaAla 508  
 Db 1297 AGTGAATCTGTCGAGGAGCTCTCATCTGATGACGTGCTGAGCAATGAAAAA 1356  
 Qy 509 GlnAsnProAsnAsnGluVal-----ThrTyrLeuLeuProGlnAlaIle 523  
 Db 1357 CAGAACAGGGCTGACGGGATTGTGAGCCCGGATTAATCATTTCTGTTGCGCGATTC 1416

Qy 524 GlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAla 543  
 Db 1417 GCGAGGGAGAGCCATTCATCGCTCTATGAAGTGGCCACCGCTGTGGTGGCGCA 1476  
 Qy 544 PheAlaThrValLeuLysAlaLeuIleGly-----LeuAspArgGlyGlyGlu 559  
 Db 1477 TGTGTGACAACTCTCAAGCGGTTCTTCGACGCCAACCTTCAGATCGATAAGGTGTCGAG 1536  
 Qy 560 CysPheProAsnProValPheProSerAspAspGlyLeuGluLeuIleAsnPheGluGly 579  
 Db 1537 -----GTCGACACTGATGAGCAAGCTTCTGAAGTCTGTTTCAAGGGA 1591  
 Qy 580 AlaCysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArg 599  
 Db 1582 ACT---CTCACTGTTCGGTGAATGAACAGCTGCCACATGTTGGATCGGCGG 1638  
 Qy 600 GlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThr 619  
 Db 1639 AACATGGCGGTGTTCACTACTTCTCTGACCACTTCGAGTCACTGCTACTCGGTGAACAG 1698  
 Qy 620 IleThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluAlaThrPheGlu 639  
 Db 1699 ATTGCGATTGGAATCTTGGAGGACGAGCTGACGTATCGCAGAACTTCTTCTTCAAT 1758  
 Qy 640 PheArgLeuPheThrGluGluValIleLysLeu 650  
 Db 1759 TTCCCGAATTCAGCACTACATCCAGATT 1791

## RESULT 13

BX294153 LOCUS 293350 bp DNA linear BCT 11-JUL-2003  
 DEFINITION Pirellula sp. strain 1 complete genome; segment 21/24.

BX294153 BX119912  
 ACCESSION

BX294153.1 GI:32447383  
 VERSION

complete genome.

Pirellula sp. 1  
 SOURCE

ORGANISM

Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;  
 Planctomycetaceae; Pirellula.

## REFERENCE

## AUTHORS

Gloeckner, F.O., Kube, M., Bauer, M., Teeling, H., Lombardot, T.,  
 Ludwig, W., Gade, D., Beck, A., Borzym, K., Heitmann, K., Rabus, R.,  
 Schleesner, H., Amann, R. and Reinhardt, R.

Complete genome sequence of the marine planctomycete Pirellula sp.

strain 1

Proc. Natl. Acad. Sci. U.S.A. 100 (14), 8298-8303 (2003)

MEDLINE 22735913

PUBMED 12835416

REFERENCE 2 (bases 1 to 293350)

## AUTHORS

Kube, M., Borzym, K., Heitmann, K., Klages, S., Marquardt, I.,  
 Lehmack, S., Beck, A., Pawlik, R., Reinhardt, R., Gloeckner, F.O.,  
 Bauer, M., Teeling, H., Lombardot, T., Ludwig, W., Gade, D., Rabus, R.,  
 Schleesner, H. and Amann, R.

Direct Submission

Submitted (21-JAN-2003) Max Planck Institute for Molecular

Genetics, proScience Ihnestrasse 73, D-14195 Berlin, Germany Max

Planck Institute for Marine Microbiology, Bremen, Germany; in the

framework of the REGX-project, http://www.regx.de

Genome Center: Max Planck Institute for Molecular Genetics

Center code: MPIMG

----- Summary Statistics

Sequencing vector: pUC19; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 7142841 bases at least Q40

Consensus quality: 7145138 bases at least Q30

Consensus quality: 7145484 bases at least Q20

-----

Center: Max Planck Institute for Molecular Genetics

Center code: MPIMG

----- Summary Statistics

Sequencing vector: pUC19; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 7142841 bases at least Q40

Consensus quality: 7145138 bases at least Q30

Consensus quality: 7145484 bases at least Q20

Quality coverage: 8.03

-----  
This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid sequence; assembly was additionally confirmed by long range PCR and cosmid end sequences.  
-----

See <http://www.micro-genomes.mpg.de/pirellula/> for more information including minimal tiling path from a set of 220 cosmids out of 908. See the misc\_feature tag below for the boundaries of the MTP cosmids. ----- Annotation

Center: Max Planck Institute for Marine Microbiology  
Celsiusstrasse 1, D-28359 Bremen, Germany.

Center Code: MPIWM

Email: fogempi-bremen.de

Phone: +49 (0)421 2028 938 Fax: +49 (0)421 2028 580

-----  
Three different programs (Glimmer, Critica and Orpheus) were used for ORF-prediction. A nonredundant list of ORFs was generated by suitable parsing of the results.

Automated annotation was done with the software package Pedant Pro (<http://www.bionax.de>). All ORF predictions and annotations were manually corrected by considering all results of the different tools applied. See <http://www.regx.de> for more information and access to supplementary information.  
-----

#### FEATURES source

Location/Qualifiers  
1. .293350

/organism="Pirellula sp. 1"

/mol\_type="genomic DNA"

/strain="1"

/db\_xref="taxon:243090"

/complement(93..1040)

/gene="uspA"

/locus tag="RB11183"

/complement(93..1040)

/gene="uspA"

/locus tag="RB11183"

/notes="PMID: 9371463 best DB hits: BLAST: pir.A69220; conserved hypothetical protein MTH898 - Methanobacterium; E=1e-07 gb:AA046412.1; AF096262.1 (AF096262) ER6 protein (Lycopersicon; E=1e-05 gb:AA041173.1; AB006246.3 (AB006246) conserved hypothetical; E=2e-05 COG: MTH898;

COG0589 Universal stress protein UsppA and related; E=1e-08 PFAM: PF00582; Universal stress protein family; E=4e-11"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="CAD78973.1"

/db\_xref="GI:32447384"

/translation="MRVLLAVDSPPYQQAVFASHLPRLKRPVDFLVSVVAPPMLVD TGGMSMPDFGSGFLEITDRSRAIDAVSLKSDQHVSHVTHVPVPGPTSLALLDVA DESADILVIGALGHSIAIERVLGVSVDYVATADMTSLVVRPTSEADVPDLQKIML ALSGRPDEMLTWRELKLPNVEHLVRLDRFSYVRDLRQQASDAWAOEAOAQ AQLIDFTKLQQLGTLTETHFVSNHVGTELVYARHGGLAVTGSDSGLLTRVPL GSTRVYLRADCSVLIIRDREDAKARQIAQSQSLAST"

/complement(1116..1241)

/locus tag="RB11184"

/complement(1116..1241)

/locus tag="RB11184"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein-signal peptide prediction"

/protein\_id="CAD78974.1"

/db\_xref="GI:32447385"

/translation="MFASRWTVRAASLTASPTCATPSHLARHSQQNASWSEPK"

/complement(1231..2142)

/gene="uspA"

/locus tag="RB11185"

/complement(1231..2142)

/gene="uspA"  
/locus tag="RB11185"  
/notes="PMID: 8905231 best DB hits: BLAST: Swissprot:P74148; YD88 SYN3 HYPOTHETICAL 17.3 KD PROTEIN SL11388; E=2e-05 gb:AAG20091.1; (AE005088) Vng1898c [Halobacterium sp. NRC-1]; E=5e-05 swissprot:Q57951; Y531 METUA HYPOTHETICAL PROTEIN M30531 -----; E=7e-04 COG: SL11388; COG0589 Universal stress protein UsppA and related; E=2e-06 PFAM: PF00582; Universal stress protein family; E=0.04"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein-putative universal stress protein"  
/protein\_id="CAD78975.1"  
/db\_xref="GI:32447386"  
/translation="MNERNGPKLRLTIGDSSAPARDALQALSKSTLAACEVSLATI VPEPLYTIDDTSMPIPOEVFDSRLLENTOLKQOEFGERFSSCFVSQGHGPR GLIEMSERFDADWIVIGSVGHSAFSLRLGSTDYVANRSTRCLVHRPITTHPEPSK GFISRVIAISNAESTALPDVAAALRPNCEVHLVNMVTHPEYAEHLKRVAAAYM SEYRSAAWKLMTETRPRLAELGMMKVPFSLLESHPVGRVAVLEYANEHACDLIVVGDDDD SLMERVMLGVSFRFVVRHANQSVLIIVR"  
/complement(2139..3614)  
/gene="sasR"  
/locus tag="RB11186"  
/locus tag="RB11186"  
/complement(2139..3614)  
/genes="sasR"  
/locus tag="RB11186"  
/function="two-component signal transduction"  
/notes="best DB hits: BLAST: gb:AAG45333.1; AF273214\_1 (AF273214) SasR [Myxococcus xanthus]; E=3e-83 Swissprot:Q06065; ATOC ECOLI ACETOACETATE METABOLISM REGULATORY; E=1e-81 gb:AAC28085.1; (AF100457) response regulator [Myxococcus xanthus]; E=5e-80 COG: atcC; COG2204 AAA superfamily ATPases with N-terminal receiver; E=1e-82 PFAM: PF00072; Response regulator receiver doma; E=2.7e-32 PF00158; Sigma-54 interaction domain; E=1.1e-119"  
/codon\_start=1  
/transl\_table=11  
/product="NtrC-like response regulator SasR"  
/protein\_id="CAD78976.1"  
/db\_xref="GI:32447387"  
/translation="MIGITMSQTAIKILLVDDDDVRRSCGKPMRWGHTVRLAASGA EAMSFIDREAFDVAFIDIMPMSGELMQRVHEESHVDFVLTGKGTIEACVQAMQ LGSADFLTPCSLADLEHRLVQLAQGHKQKQKAIYVHRNRPSPKLGSLNLSME LGSMIAKVAPTKRPVLIEGSGTGKEVARSVOQSLADRPFTVINCAALPAOLVES ELFGHCKGSGTGAADQGLFEVADGGLFIDVSGELPPALQPLKLVLDGSGLRRLG CHRRKVKVRLIAATNRDLLEEVKAGRFREDLLYRINVLINLPLPQRSGDIELLLH HFLPRGWSFDKEALDTMVQVHWPGNVRLINVERATILADHQVITMDDLPRGSLNLP KVLQELGKNKAKTARHLGHRRLYLLERFEQQMPAETL"  
/complement(3740..4012)  
/locus tag="RB11190"  
/complement(3740..4012)  
/locus tag="RB11190"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="CAD78977.1"  
/db\_xref="GI:32447388"  
/translation="MIEISHGARRQRNGNITSPFVLKTRASITRSHRERHMLLFFV TPPLNPTCPNSRPLHRHSIRFDADVVTGLTDFRAGCFHAPRPAPV"  
/complement(3980..5446)  
/genes="hydH"  
/locus tag="RB11191"  
/locus tag="RB11191"  
/complement(3980..5446)  
/genes="hydH"  
/locus tag="RB11191"  
/EC\_number="2.7.3.-"  
/function="two-component signal transduction"  
/notes="PMID: 2666400 PMID: 11243806 best DB hits: BLAST: gb:AAG59806.1; AF305914\_1 (AF305914) HydH [Klebsiella oxytoca]; E=2e-23 gb:AAG09417.1; AF180147\_9 (AF180147)

bZIP histidine kinase Tobs; E=5e-23 embi:CA43735.1;  
(Y18245) bZIP histidine kinase [pseudomonas; E=6e-23 COG:  
BH1920; COG0642 Sensory transduction histidine kinases;  
E=2e-22 PFAM: PF00072; Response regulator receiver doma;  
E=1.2e-21 PF00989; PAS domain; E=1e-08 PF00785; PAC motif;  
E=0.053"  
/codon\_start=1  
/transl\_table=11  
/product="sensor kinase of the HydH/G two-component  
system"  
/protein\_id="CAD78978.1"  
/db\_xref="GI:32447389"

## Alignment Scores:

Pred. No.: 1.35e+03 Length: 293350  
Score: 135.50 Matches: 172  
Percent Similarity: 30.98% Conservative: 82  
Best Local Similarity: 20.98% Mismatches: 295  
Query Match: 3.04% Indels: 272  
DB: 1 Gaps: 38

US-10-691-383-2 (1-676) x BX294153 (1-293350)

Qy	7	AspThrThrArgGlySerProMetProAspThrGly	-----ValLeuArgLeu 22
Db	275396	GACAGAAATCCCGGACCGGAGGCAACACACAGCGCGCACCTGCCATTG	275455
Qy	23	LeuThrSerGluGlnArgAlaLysGlyTrpArgArgGlnLeuGluGlyGlyLysSerLeu 42	
Db	275456	TTATTGTGAT-TCTCGATGACATGGCTTCAGAAAGG-----TCCCGA	275496
Qy	43	GlyPheHisProSerGlu-ThrProTyrIleLysTyrLeuGluGlySerGluThrTrpLy 62	
Db	275497	GGACTTCACCTCAACCATACCGCGATCGCGCGAGTACCAGTGGTGCACCGGTGCGGT	275556
Qy	62	sLysValLysLeuProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIleMetAl 82	
Db	275557	GGACACGGTCACTCCACCGCACGCGTACGACGACG-----	275593
Qy	82	aArgValArgIleAlaThrAlaLeuAlaValValLeuAlaProCysLeuAlaPheAs 102	
Db	275594	-----ATCGTCGACGA	275604
Qy	102	pGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGlyGlyArgHi 122	
Db	275605	TGATTGACCC-----AACTGGCAGTT	275625
Qy	122	sLeuGlnThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArgArgAs 142	
Db	275626	GTTCGACACC-----GGACCGACTACATCGATGAGGCTGA	275661
Qy	142	pAsnValAlaPheAlaSerArgArgAlaAlaArgArgGluArgAspGlyThrGlyTh 162	
Db	275662	TTCCGGCATGTACTCGTGGGTCTGACCGGTGCTGCGGTGAGCGCTCGCGAGACGAC	275721
Qy	162	rValCysGlnIle---ThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSerLe 181	
Db	275722	GATTGACATTGTCGTACGCTCAACTCCCAATGACGGCCCGGACTTTACGGAGACATT	275781
Qy	181	u-----	181
Db	275782	GGGTGAGCAATTCAAGAAGTGTGCGAGCTTACAATGCGCGCGCGGTGCGGCCCCCT	275841
Qy	182	-----ProHisAspGlu-----LeuGlyGlnValThrAlaAs 192	
Db	275842	TGCAAGATCACTCGCGACGATAATTCGTTTACATACATACTCGATGGGAACCGCTTCAC	275901
Qy	192	pAspPheAlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAspVa 212	
Db	275902	CGGTGACGGATTGGCTAACACCGCGCGCTTGCATCGACTGACGACGTTTCGATGATAC	275961
Qy	212	lProAlaGlyAspProAla-----GlyArgLeuValAsnProThrAlaAlaPheAlaI 230	

Db	275962	GAACGCAAGTCAACCGCGGCACCTGGCCAAATTTGGTTGAACCCACAGAGTCTTTCAGCAT	276021
Qy	230	eAspIleSerGlyProAlaPheSer-----AlaTh 240	
Db	276022	CGACATCAGCAATCCAGGCGGATCGAATGTTCCCGGATCGGAATTTACAGTCCCGGAC	276081
Qy	240	rThrIleProProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLe 260	
Db	276082	GCAAGCGACCCCGGTGACTGACGATGACTGCTGTAATTCACATCGAAGCAATCCAG--	276139
Qy	260	uTyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGluIleTh 280	
Db	276140	-----CAGCGCTCGGAGGATCGG-----GTTGGGTCCAGGACCCAGCGAAATCTT	276183
Qy	280	rThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSe 300	
Db	276184	CAGAGTTTCGATGTCACAC-----CCGAGCCAGAAATAGTGTCTC	276222
Qy	300	rIleGlySerAspGlyThrValAspProPheSer-----GlnLeuPheArgAl 316	
Db	276223	GGTCAACTTCACCTGACCAAGTTGTTCCACAGACAGCTCAGCCGCGCAAGATTACCAAGC	276282
Qy	316	a-----ThrPheValGlyValGluThrGlyProPheValSerGlnLe 330	
Db	276283	CATCCTTGGCTCGGCGGTATTGTTCACTGGCAATTCGACGCTTCCA-----AA	276330
Qy	330	uLeuValAsnSerPheThrIleAspAlaIleThrValGluProLysGlnGluThrPheAl 350	
Db	276331	GCTGATCAGTGTGATACGATGACGACTGTTGGTTCGAAGGCGATGAGTCCGTCATTGT	276390
Qy	350	aProAspLeuAsnTyr-----MetValAspPheAspGluTrpLeuAsnIleGl 366	
Db	276391	GAATTTGACCGAATTCACITTCACCGGATGGTTGACCCGGATACAGATT---GCTGC	276447
Qy	366	nAsnGlyGlyProProAlaGlyProGluGluLeuAspGluGluArgPheIleArgAs 386	
Db	276448	AACGGGTGGTACCCCGCACCGACCGCGACACTGCTGCTGTCGATCAGCATCACCGCAA	276507
Qy	386	nAlaArgAspLeuAlaArgValSerPhe-----	395
Db	276508	C-----GACTTAGCTGAAGTTGAATTTCACTGGCGCAACACAGACACAGACGACGAC	276561
Qy	396	-----ValAsnIleAsnThrGluAlaTyrArgGlySerLe 408	
Db	276562	TGCCCTCGACATCACTGTTGAGTTGAATATGGTGGGTACATCTACCGCTTTAGCACCGC	276621
Qy	408	uIleLeuLeuGluLeu-----GlyAlaPheSerArgProGly----- 420	
Db	276622	AGTCACTGTCGATGATCTGATCTGACGCGGAAGTGCAGATTGATCCCGCAGATTACAG	276681
Qy	421	-----IleAsnGlyProPheIleAspSerAspArgGln----- 431	
Db	276682	CCTGAGTGCACCGTGTCACTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	276741
Qy	431	----- 431	
Db	276742	CACACTCACACCTGTTGATGACAGCTTGGTCGAGGGTGCAGAAACCGTCGACCTTGAGTT	276801
Qy	432	-----AlaGlyPheValAsnPheGly---ThrSerHisTyrPheAr 444	
Db	276802	GCAAAACCTGACACACATGAGTGGCCAAAGTTTCGCTGGGTACGACACCGCATCAGACGGT	276861
Qy	444	gLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnValHi 464	
Db	276862	CACATCACCGCAATGAGTGGCTGAG-----TGAATCTAGT	276900
Qy	464	sArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIle----- 480	
Db	276901	TCAAAGTGGATCGCTGTGACCGAGGTGGCACGGTGATACACGTTCTCGCTGACCGG	276960
Qy	481	-----AlaGly---AspLeuAspAlaAspPh 488	
Db	276961	TAGCGGACGCGGTGGGCGACCTTACTTGTTCGAAGCGGAGAGATCATCGGTGCGACGT	277020



```
misc_feature
664..665
/notes="cosmid pircos-c3h04/ cosmid pircos-blf02 joining
point"
1740..2954
/locus_tag="RB1856"
CDS
1740..2954
/locus_tag="RB1856"
/EC_number="4.2.1.24"
/functions="biosynthesis of vitamins, cofactors, and
prothetic groups"
/notes="PMID: 8226669 best DB hits: BLAST: ddbj:BA35069.1;
(AB015492) porphobilinogen synthase; E=1e-114
swissprot:P45622; HEM2 BRAJA DELTA-AMINOLEVULINIC ACID
DEHYDRATASE; E=1e-108 wisprot:Q59643; HEM2 PSEAE
DELTA-AMINOLEVULINIC ACID DEHYDRATASE; E=1e-105 COG:
PA5243; COG0113 Delta-aminolevulinic acid dehydratase;
E=1e-106 PFAM: PF00490; Delta-aminolevulinic acid
dehydratase; E=1.8e-183"
/codon_start=1
/transl_table=11
/product="delta-aminolevulinic acid dehydratase"
/protein_id="CAD72301.1"
/db_xref="GI:32396997"
/translations="MGWVGRGEIRKPPFNVIIRGILSAERPGRLGTFCDAKWKHLAA
DECPKITSAPVLSILWVLSQSGPAPRAFPATRLRRVRATOWSRRLVRETTLS
VDDLWPLFVMDSGSQPQSLGVNRLGESEIVAAKRAVDLGI PAIALFPATDPKL
KSDAAEAYNSDLNCRVTRQIKDAVGDLSGLVLDVALDPYSHGGQGLVQDQVIND
ETVDVLCQAIQVAAAGCDVIAFSDMMWDRIGRAISALDGAHSGVQVMSYAAKYASA
FYGFPRDAGNLAAGAAKKTKYQSPQSDEAIAEVALDLAGADSVNVKPMQPYLD
IVARVQTFCVPTTFAYQVSGEYAMLRGAADAGWLSGDAVILESLLSFKRAGADVITY
FAADAAELLHRS"
2980..3774
/locus_tag="RB1858"
CDS
2980..3774
/locus_tag="RB1858"
/functions="protein modification (glycosylation,
acylation, myristylation, palmitoylation, farnesylation and
processing)"
/notes="PMID: 12093901 best DB hits: BLAST: pir:H82553;
3-demethylubiquinone-9 3-methyltransferase XF2471; E=0.027
pir:T30158; hypothetical protein C37A2.6 - Caenorhabditis
elegans; E=0.030 embi:CAA89660.1; (Z49629) ORF YJR129c
[Saccharomyces cerevisiae]; E=0.76 COG: XF2471; COG2227;
E=0.003"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD72301.1"
/db_xref="GI:32396998"
/translations="MVQDRPAHSASRSSTPMNATVLEDHETGDKTLLFSELSQT
QFSMADSLFAWHDTLIAGQERPLAVASDPDAMLIACERDAGEGVDPFWATT
WRAAGLDRELDVPIHQNDLVRCVGGTGAGIALLRGARVLTLDGVDPQLVRLS
LRLGLHADVQVFLGDESLAPKPPFLLGSDVTLRLTWPELLOAAEHLTDGQLI
LSDPQRLIAIEFSQVWKDKPDVTEHTVEMDDDDPEHPIRIMVLTQG"
3832..4206
/locus_tag="RB1859"
CDS
3832..4206
/locus_tag="RB1859"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD72302.1"
/db_xref="GI:32396999"
/translations="MSDFASRLDPNGSEGSYESSNVLRFPSPKQSRVTVVPNCLRVD
GSIESREANAVNLRSGVSHDVLIRGGTIGGNCSTLVKARACIALFQSIGSSVRYNLT
YAVTHAPKQSRPGSTWRRLDGG"
4261..4593
/locus_tag="RB1860"
CDS
4261..4593
/locus_tag="RB1860"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD72303.1"
/db_xref="GI:32397000"
/translations="MSVRYNLTYSKPAQVPPGGGWMVNTALASRLILASLCSKAS
SWPPGTTWPPGCPAKTKSKSLDPTTYLLYRLTSTPVNQLMASVRTCSLQSTLPTAFR
SILRSFAK"
gene
4530..4937
/locus_tag="RB1862"
CDS
4530..4937
/locus_tag="RB1862"
/EC_number="5.4.1.16"
/functions="transcriptional control"
/notes="PMID: 9384377 best DB hits: BLAST: pir:G69999;
transcription regulator GntR related protein ytrA -;
E=8e-17 pir:D72336; transcription regulator, gntR family -
Thermotoga; E=2e-14 pir:T36493; probable gntR-family
transcription regulator -; E=9e-11 COG: BS ytrA; COG1725
Predicted transcriptional regulators; E=8e-18 BS ytrA;
COG2188 Transcriptional regulators; E=8e-08 BH1164;
COG1725 Predicted transcriptional regulators; E=8e-08
PFAM: PF00325; Bacterial regulatory proteins, crip f;
E=0.49 PF00392; Bacterial regulatory proteins, gntR;
E=9.4e-16"
Alignment Scores:
Pred. No.: 2 06e+03 Length: 259950
Score: 132.00 Matches: 162
Percent Similarity: 32.72% Conservative: 88
Best Local Similarity: 21.20% Mismatches: 301
Query Match: 3.74% Indels: 216
DB: 1 Gaps: 32
US-10-691-383-2 (1-676) x BX294136 (1-259950)
QY 30 LysGlyTTPArgArgGlnLeuGluGlySerLeuGlyPheHisProSerGluThr 49
Db 59781 CGCGGGTGAACCAACGACATGCAGGC-----TTTGGCGGATGTCTGGAACCAAT 58831
QY 50 ProTyrIleGlySerLeuGluGlySerGluThrTyrLysLysValLysLeuProThrAsp 69
Db 58832 CA-TCGCTCACTATGTTGATGGCGATGACGACATCGCGGAAACGCGGATGATGTCACA 58890
QY 70 GlyIleSerAlaSerLysIleLeuGlyLysIleMetAlaArgValArg-----85
Db 58891 AGGTCACTTGTTCAGCCATCTACAGCAGATTTTCGTTGGAGGAAGATTCGCTGGATTGC 58950
QY 86 -----IleAlaThrAlaLeuAlaValLeuAlaAlaProCys-----98
Db 58951 TCAACGGTGTGCGACGTCATTCGCGCGTGTTCGGACACGCTGCTCGCGATGCGCAAG 59010
QY 99 ---LeuAlaPheAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThr 117
Db 59011 ACACITTTGCGTGCAGCGGATGTA-GCCGAGGAAATATTCCT-----59051
QY 118 GlyGluGlyArgHisLeuGlnThrCysThrAsnSerAspAlaLeuAspProThrAla 137
Db 59052 -----TTGCNAACG---ACCAACGCGGACAAAC-----GATCCGACGCTG 59087
QY 138 ProAsnArgArgAsp-----142
Db 59088 ATCGTCAGCACCGCAGCGGAATACAGTTCGCGCGGTTGGTTCGCTCGACTTTTGACGCC 59147
QY 143 ---AsnValAlaPheAlaSerArgArgAspAlaAlaArgArg-----GluArgAsp 158
Db 59148 AGTGGTGTGCTGATTCATCATGATGTTGGACGCTTCGCTACGCGAGTCTTTGGCCACTGAC 59207
QY 159 GlyThrGlyThrValCysGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPheHis 178
Db 59208 GAAGCGGGAACATGTG-----GTCGTCACCTGGTGGCGCAGACATTCGCCCGGATCGCT 59261
QY 179 LysSerLeuProHisAspGluLeu-----GlyGlnValThr 190
```

59262 GATAGTACCAAGCGGACCGAAGTCAACAACTGGTTCAAGCGGTTCGGGTGTTGTGCTACT 59322  
Qy 191 AlaAspAspPheAlaIleLeu-----GluAspCysIleLeuAsnGlyAspPheSerIle 208  
Db 59322 GCCAAAGACGGAGAGATCTTTGGTGCACCGATGTCTTCTCGAAGGACGCGCGCAGAG 59381  
Qy 209 CysGlu-----AspValProAlaGlyAspProAlaGly 219  
Db 59382 GTTTCGAACACAAGCAACGAATCTGGGCAACCTGCACGCGGACGCGCAACCTGTGTGCAGCT 59441  
Qy 220 ArgLeuValAsnProThrAlaAlaPheAlaIleAspIleSerGly----- 234  
Db 59442 CGTCAGTTGCATCTACAGTGGCGGTTTCGATCAAAATATGGTGGCGGATCGTCGCGCGCG 59501  
Qy 235 ---ProAlaPheSerAlaThrThrIleProProValProThrLeuSerSerProGluLeu 253  
Db 59502 ATTGGTTCGATCGATCGATCGAGCGGTGAAGACTTCCGACCGCGCGCAACCCAGCGCA 59561  
Qy 254 AlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAlaArgAspValProPheMetGln 273  
Db 59562 GGAAAGGAACGACGGCGAGATCTCTCAACTGGATATCGAAACCGCTTGGCGGTTT---GAC 59618  
Qy 274 TyrGlyThrAspGluIleThrThrThr----- 282  
Db 59619 AACGGTTTGACCGTGATGACGGTCAACCGGATGGAGCTGAACGAGTTGATCGAACACGGC 59678  
Qy 283 ---AlaAlaAlaAsnLeuAlaGly-----MetGlyGly 292  
Db 59679 GTTTCGCGGACCGATACGCGCGGCAACGACGACGACGAGTTCAGTTCGCCCAAAATCGGTGT 59738  
Qy 293 Phe-----ProLeuLeuAspAlaValSerIleGlySerAspGlyThrVal 307  
Db 59739 GCCAAGTTCAGCTTTGATCCAAACGAACCAACGACATTGGTGTGGACGCTTCGCCCAACGTC 59798  
Qy 308 AspProPheSerGlnLeuPheArgAlaThrPheValGlyValGluThrGlyProPheVal 327  
Db 59799 GTTCAGAAAGCCAGCGAGTTCGACGCTTCGATTTGGTTCGACGTAACGGTGAATATCATC 59858  
Qy 328 SerGlnLeuLeuValAsnSerPheThrIleAspAlaIleThrValGluProLysGlnGlu 347  
Db 59859 GACGCTCTTGGTGCAAATGGCCAAATGATCGGC-----AATCCCAANTCGCGAA 59906  
Qy 348 ThrPheAlaProAspLeuAsnTyrMetValAspPheAspGluTrpLeuAsnIleGlnAsn 367  
Db 59907 ATTTCGATATCGTACTTTGGACTTCCTCGCT-----GGGATCTTCTCG 59948  
Qy 368 GlyGlyProProAlaGlyProGluGluLeuAspGluGluLeuArgPheIleArgAsnAla 387  
Db 59949 GCGCGTTCACCGATTGGAGGC----- 59969  
Qy 388 ArgAspLeuAlaArgValSerPheValAspAsnIleAsnThrGluAlaTyrArgGlySer 407  
Db 59970 -----GATGGCTACCCGTTCCCGCCCTTCGGTGATTCC 60002  
Qy 408 LeuIleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAsp 427  
Db 60003 ATGACTCATCTTTTCGTCGGCTGTTGTGCG---GATGGTGCAGCGAGTGTCTTTCACGACC 60059  
Qy 428 SerAspArgGlnAlaGlyPheValAsnPheGlyThrSerHisTyr-PheArgLeuIleGln 447  
Db 60060 GGTAAAGGACGAGGATCGCTGGTGGTGAATACCTGCTGGCCCAACCAACATCCGGATGGTGGG 60119  
Qy 447 y-----AlaAlaGluLeu-AlaGlnArgAlaSerCysTyrGlnLysTrpG 462  
Db 60120 ACCGCTCTTATCGAGAGACGAGACTTCGCCAGTTCAAGATCAACGATTCACAAATTTG 60179  
Qy 462 InValHisArgPheAlaArgProGluAla---LeuGlyGlyThrLeuHisAsn-ThrIle 480  
Db 60180 GCGGTCAACCAATGACACGATCTGACGCCATCTCTCGATCGACAGTTTCAAAATCTCCGTT 60239  
Qy 481 AlaGlyAspLeuAspAlaAsp---PheAspIleSerLeuLeuGlu-----AsnAsp 496  
Db 60240 GCCCGCACTTCGAAACGGCGGTGTTCAGCAAGTATCGCGCGCAATCGTGCTCATGAC 60299

Qy	497	GluleuLeuLysArgValAlaGluleuleuLeuAlaAlaGlnAsnProAsnAsnGluValThr	516
Db	60300	CCGCAAAACCCAGCGAGTCTTCTTACGAACTCCGATGATGCAACACCAITGGT	60350
Qy	517	TyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisPro-	531
Db	60351	-----GTGCTGGACATTCGCCGATCCCAACTCACCAACGGAACCTTCCCCC	60395
Qy	532	---SerTyrProSerGlyHis-----AlaThrGlnAsnGlyAlaPhe	544
Db	60396	ATCACGTTTCCACGCGGACCGGTGGTGTGAACAGCGTTCGCGGTTCGTGGATCATGTT	60455
Qy	545	AlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCysPheProAsnPro	564
Db	60456	GCAGTCGCGGTGCTCTCCAACGCACACTGATCTCTGTGGAGTTCTGTTC-----	60506
Qy	565	ValPheProSerAspGlyLeu-GluLeuIleAsnPheGluGlyAlaCysLeuThrTy	584
Db	60507	---TTTGACGTGACCGGCAACTTACTCGGCAGTGTACCGTTGGGGCTCTGCCAGACTCA	60563
Qy	584	rGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIle	604
Db	60564	CTCA-----	60567
Qy	604	eHisTyrArgPheAspGlyIleGlnGlyLeuLeuLeuGlyGlu-----	618
Db	60568	-CAITTTTCTCCC-GACGGGATGAAAGTGTGCTTCTGTGTGAAGCGCAACCGAGATTT	60625
Qy	619	-----ThrIleThrVal-----Ar	623
Db	60626	GGAGATCCCAATCCAGCATCCCTTGGGTACGATCAGCGTGAATGATCTGACTCG	60685
Qy	623	gThrLeuHisGlnGluLeuMetThrPheAlaGluAlaThrPheGluPheArgLeuPh	643
Db	60686	ATTGCGTCAAGCGGTTTCATCACAATCTCTTCAGCTGACCGCTCGACTTCACGCGGTT	60745
Qy	643	eThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIleAspGlyAspMetCy	663
Db	60746	TGATGACAGAAGACTCGCTACAGGTGTACGATCTTCCCGGCGCGCTCGGC	60805
Qy	663	sSer 664	
Db	60806	GTCA 60809	
RESULT 15			
AX598603		9579 bp	linear
LOCUS			
DEFINITION	Sequence 11 from Patent WO02088176.		PAT 14-FEB-2003
ACCESSION	AX598603		
VERSION	AX598603.1	GI:28398756	
KEYWORDS			
SOURCE	Streptomyces platensis		
ORGANISM	Streptomyces platensis		
REFERENCE	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
AUTHORS	Streptomycineae; Streptomycetaceae; Streptomyces.		
TITLE	1		
JOURNAL	Farnet,C.M., Zazopoulos,E., Staffa,A. and Yang,X.		
FEATURES	Genes and proteins for the biosynthesis of polyketides		
source	Patent: WO 02088176-A 11 07-NOV-2002;		
	Ecopia Biosciences Inc. (CA)		
	Location/Qualifiers		
	1..9579		
	/organism="Streptomyces platensis"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:58346"		
	/notes="subsp. rosaceus"		
ORIGIN			
Alignment Scores:			
Pred. No.:	27	Length:	9579
Score:	131.00	Matches:	152
Percent Similarity:	30.12%	Conservative:	82

Best Local Similarity: 19.56%		Mismatches: 254	
Query Match: 3.71%		Indels: 289	
DB: 6		Gaps: 38	
US-10-691-383-2 (1-676) x AX598603 (1-9579)			
Qy	5	AlaAlaAspThrThrArgGly---SerProMetPro-AspThrGlyValLeuArgLeuLe 23	
Db	6726	GCCGACATGAGCAGCGCGCCCGACCTCGACCGGGAGCACTTCTACGCGCGCTCGG 6785	
Qy	23	uThrSerGluGlnArgAlaLysGlyTrpArgGlnLeuGluGlyLysSerLeuGl 43	
Db	6786	CACCGCGGCTCGACGCGCGCGCTACGCGCGCAGCGTCAAGGGTCAACGAACTCGA 6845	
Qy	43	yPheHis-----ProSer-----GluThrProTyrIleLy 53	
Db	6846	CGCCACCGGCTGCTCGTACGGGTGCGCGAACCGGCCATGTGCGAGGACCGCAACAGCA 6905	
Qy	53	sTyrLeuGluGlySerGluThrTrpLysLysValLysLeuProThrAspGlyIleSer-- 72	
Db	6906	GCAGTGCAT---CTCCGCGCTGGCGCTGCTCGGGCTG---ACCAGGGTGTTCAGCA 6959	
Qy	73	-----AlaSerLysIleLeuGlyLysIleMe 81	
Db	6960	CGCGTGGGCGGCGCGACGCGCGCGTGTGCGGTTCGATCCGTGCGAGGCGAGCAGTG 7019	
Qy	81	tAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPh 101	
Db	7020	GGAGCGCACCCGG-----GCGATGCTGCTGGCGCGG----- 7050	
Qy	101	eAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGly----- 118	
Db	7051	-----ACGTCCGACGCCGTCTC-----CATCGGCTTCTCTCGA 7085	
Qy	119	----GluGlyArgHisLeuGlnThrCysThrAsnSerAsp-----AspAlaLeuAs 134	
Db	7086	CGAGGACGGCGGTGCTGGCGGGTGGAGGACCGCGAGTTTACCGCGGCGCACCTTGA 7145	
Qy	134	pProThrAlaProAsnArgArgAspAsnValAlaPheAla----- 147	
Db	7146	GCCGGCACTCCCGGTGAGCGCGGACCGGCACCTCGTGGCACCTGCGCAGGCGTGGCGTCC 7205	
Qy	148	-----SerArgArgAspAlaAlaArgArgGl 156	
Db	7206	GTCGCTGGAGACGCGGTGTGTACGGGGAGTGGCAGCAGTGGAGGCGGTGGCGCGGA 7265	
Qy	156	uArgAspGlyThrGlyThrValCysGlnIleThrAsnGly----- 169	
Db	7266	GGCGAGCGCTCGTGACCGTGTGCGCGGTGCGGACGCGGCGCGGCGGCTCGTGGCGTC 7325	
Qy	170	-----GluThrAspLeuAlaTh 175	
Db	7326	GTCGCGGAGACGTCGCGCACCTGCTCAAGTTCGACTGCGGGACATCGACCTCGACAC 7385	
Qy	175	rMetPheHisLys---SerLeuProHisAspGluLeuGlyGlnValThrAlaAspAspPh 194	
Db	7386	GCACTTCCACGCTACGGCTTCGAGTCCATCGCGCTGGCCAACTGGCCTCGGAACCTCAA 7445	
Qy	194	eAlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAspValProAl 214	
Db	7446	CGGCGTCTCGGACGCGACCTCACCCCGCGCTCTTCTTCGAGTGTCTCCGAC----- 7497	
Qy	214	aGlyAspProAlaGlyArgLeuValAsnProThrAlaAlaPheAlaIleAspIleSerGl 234	
Db	7498	-----ATCCGACGCTCGCGGAGTACTGTCTCGACCGCTACCGCTACGG 7535	
Qy	234	yProAlaPheSerAlaThrThr-----IleProValProThrLeuSe 249	
Db	7536	CCCCGAGCTGAGCTCCCCACGAGCGCGACCGCCCGCGCGCTCGCGCGCCCGCGCC 7595	
Qy	249	rSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAlaArgAspVa 269	
Db	7596	GTCCCCA-----GT 7604	
Qy	269	lProPheMetGlnTyrGlyThrAspGluIleThrThrThrAlaAlaAsnLeuAlaGl 289	
Db	7605	GCCGATGCGCGCACCCGCGCGCGGACGAC-----GACGCGTGGCCATCGTCGGCGC 7655	
Qy	289	yMetGlyGlyPheProAsnLeuAspAlaValSerIleGlySerAspGlyThrValAspPr 309	
Db	7656	TGCGGAGCGGTTCCTCCCGCGCGGAC-----GACCTGGACAC 7691	
Qy	309	oPheSerGlnLeuPheArg-----AlaThrPheValGly-----Va 321	
Db	7692	CTTCTGCGACAGCTGCGCGCGGAGGACCTGATCGCGACTACCCCGCGCACCGCTT 7751	
Qy	321	lGluThrGlyProPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIleTh 341	
Db	7752	CGACGGGGCCCTTACCGGAGGTGTGCGCGGGGACTTCCGAGTTTCCGGCGC 7811	
Qy	341	rValGluProLysGlnGlnThrPheAlaProAspLeuAsnTyrMetValAspPheAspGl 361	
Db	7812	GATCGAG---GGCGTGGACCGCTTCGACGCGGACTTCTTCCACTGTGCGGCTGGAGGC 7868	
Qy	361	u-----TyrLeuAsnIleGlnAs 367	
Db	7869	GGAGCTGATGGACCCCGCAGCACCGGCTGGCCCTGGAGACCGTGTGGGCGCGCTGGAGAA 7928	
Qy	367	nGlyGly-----ProProAlaGlyProGluGluLeu-----As 378	
Db	7929	CGCGGCTACGCGCGCGCGCTCCCGAGAACACCGGGGTCTACTTCCGGGTCTCCGG 7988	
Qy	378	pGluGluLeuArgPheIleArgAsnAlaArgAspLeuAla----- 391	
Db	7989	CAGCGACTACCACTGCTGTCAAGCCAGTGGGTGGCACCGCGGTTCACCGGCCAC 8048	
Qy	392	-----ArgValSerPheValAspAsnIleAsn----- 400	
Db	8049	CGGCAACGCCCACTCGATGCTGGCCAAACCGGACTCTCTACGTCTCTGACGTGACGCGGCC 8108	
Qy	401	-----ThrGl 402	
Db	8109	GAGCGAACCGCTCGACACGCGCTGTCTCAGCTCGTCTGCGCTGACCGCGCGCTCGA 8168	
Qy	402	uAlaTyrArg-----GlySerLeuIleLeuLeuGl 412	
Db	8169	GCACATCCGTCGCGCGCGATGCGAGATGGCCATCGCGGGCGGTGTCAACTGTCTGTAG 8228	
Qy	412	uLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSerArgGlnAl 432	
Db	8229	CGTGGACACCTTCGCGCGCAGCAGCATGCGCGGC---ATGCTCAGCCCCCGCGCGCTG 8285	
Qy	432	aGlyPheValAsnPheGlyThrSerHisTyrPheArgLeuIleGly---AlaAlaGluLe 451	
Db	8286	CAAGACCTTCTCGCGCGCGCGGACCGGTACGTCCGCTCGAGGGGCTGCGCGGTGCT 8345	
Qy	451	uAlaGlnArgAlaSerCysTyrGlnLys-----TrpGlnValHisArgPh 466	
Db	8346	GCTCAAGCGCTCGCCCGCGCGCAGCGGGACGCGGCGCATCTGGGGCGTCTGTCGCGGG 8405	
Qy	466	eAlaArg----- 468	
Db	8406	CAGCGCGGAGAACACCGCGCGCGCGGTTCGCTGACCGCCCCCAACGCGAAGCGCA 8465	
Qy	469	-----ProGluAlaLeuGlyGlyThr-- 475	
Db	8466	GGCGCGCTGATCCAGGACGCCATGCGCGGCATCGACCCGGACAGCATCGGTACGTGCA 8525	
Qy	476	-LeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAs 495	
Db	8526	GGCGCACGGCACGGGACCGGC---CTGGGCGACCGCGTCAAGGTCAACCGCTCGACAG 8582	
Qy	495	nAspGluLeuLeuLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluVa 515	
Db	8583	CGCTACCTCGCGCTCGGCAACCGCGGAC----- 8610	



```
Qy      515 lThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSe 535
Db      8611 -----GGCGGCGCGCCGACGCGCGCGCGCGCGTG 8639
          ||| ||| ||| ||| |||
Qy      535 r-----GlyHisAlaThrGlnAsnGlyAlaPheAlaTh 546
          ||| ||| ||| ||| |||
Db      8640 CGCGCTCGGCTCGGTGAAGACCAACATCGGCCACGCGGAGTCGCGCGCGCGCGCTTGCGCCG 8699
          ||| ||| ||| ||| |||
Qy      546 rValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCysPhePro----- 562
          ||| ||| ||| ||| |||
Db      8700 AGTGCTGAAGGTGCTGCTGCCATG---CGTCACCGCGAGCTGCCGCGCGCGCTTGCACTG 8756
          ||| ||| ||| ||| |||
Qy      563 -----AsnProValPheProSerAspGlyLeuGluLeuIle 575
          ||| ||| ||| ||| |||
Db      8757 CGACCGGCTCAACCCGACCTGCGCTCGACGCGCGGATTGAGGTCGTA 8805
          ||| ||| ||| ||| |||
```

Search completed: September 19, 2004, 07:31:11  
Job time : 8839.83 secs

***This Page Blank (uspio)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2004, 03:16:07 ; Search time 125.3 Seconds  
(without alignments)  
1702.240 Million cell updates/sec

Title: US-10-691-383-2

Perfect score: 3528

Sequence: 1 MLCHAADTTTGRSGMPDGTGLV.....SIDGDMCSGLVYTGVDACQA 676

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3528	100.0	676	10	O82433 fucus disti
2	1131.5	32.1	646	10	Q7X9V1 laminaria d
3	1126	31.9	682	10	Q7X9V0 laminaria d
4	626	17.7	598	10	Q8LLW7 corallina o
5	625	17.7	598	10	O81959 corallina p
6	614.5	17.4	597	10	O81960 corallina p
7	133.5	3.8	8173	16	Q7UDU8 rhodopirell
8	128	3.6	706	16	Q7UWV2 rhodopirell
9	127.5	3.6	650	17	Q9HNK0 halobacteri
10	125.5	3.6	4210	16	O89J13 bradyrhizob
11	124.5	3.5	7716	16	Q7UWZ8 rhodopirell
12	123.5	3.5	959	16	O87J64 vibrio para
13	123.5	3.5	3798	2	Q9L8C6 polyanium
14	123	3.5	1939	2	Q84HP6 amycolatops
15	122.5	3.5	469	16	Q9FCJ6 streptomyce
16	120.5	3.4	6310	16	Q88PP2 pseudomonas

#### ALIGNMENTS

#### RESULT 1

```

ID 082433 PRELIMINARY; PRT; 676 AA.
AC O82433;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Vanadium bromoperoxidase.
OS Fucus distichus.
OC Eukaryota; stramenopiles; Phaeophyceae; Fucales; Fucaceae; Fucus.
OX NCBI_TaxID=3012;
RN [1]
RP SEQUENCE FROM N.A.
RA Vreeland V., Ng K., Epstein L.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF053411; AAC35279.1; -
DR HSP; P81701; IQ19.
DR GO; GO:0004601; F:peroxidase activity; IEA.
DR InterPro; IPR008934; AcPase_VanPerase.
DR InterPro; IPR000326; PA_PTPase.
KW Peroxidase.
SQ SEQUENCE 676 AA; 73344 MW; AA8A46F539A14EC7 CRC64;

```

```

Query Match 100.0%; Score 3528; DB 10; Length 676;
Best Local Similarity 100.0%; Pred. No. 2.4e-258;
Matches 676; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLCHAADTTTGRSGMPDGTGLVRLTLTSEQRAGWRRLQLEGEKSLGPHSPSETPIKYLEGSET 60
    |||
Db 1 MLCHAADTTTGRSGMPDGTGLVRLTLTSEQRAGWRRLQLEGEKSLGPHSPSETPIKYLEGSET 60
    |||

Qy 61 WKKVKLPDGTGISAKILGKIMARVRIATALAVLAAPCLAFDEVTAGVPEEHKHTGEG 120
    |||
Db 61 WKKVKLPDGTGISAKILGKIMARVRIATALAVLAAPCLAFDEVTAGVPEEHKHTGEG 120
    |||

Qy 121 RHLOCTNSDDALDPTAPNRDNNVAFASRRDARRRDGTGTCQITNGETDLATMFHKS 180
    |||
Db 121 RHLOCTNSDDALDPTAPNRDNNVAFASRRDARRRDGTGTCQITNGETDLATMFHKS 180
    |||

```



```
Qy 116 HTGERHLQCTNSDDALDPTAPNRRDNVAFASRRDAARRERDGTGTVCQITNGETD--- 172
Db 65 QPLLSGNVCRVDSFLDPVP---RAKVTL--KRLAIKAEISVSGPTCHVNGNENVP 120
Qy 173 -LATMFHKSUPHDELGOVTDADDFAILEDCTILNGDFSICEVDPAGDPAGR-----LVNPTA 226
Db 121 LFAQYHKTLPDHPKFGQVDEADYKLLLECFTSDINECEKVPFG--AGRRGGAKLTNPLG 178
Qy 227 AFAIDISGPAFSATTIPVPTLSSPELAQAELLYMALARDPVFMQYGTDEITTTAAAN 286
Db 179 GTAQVGTGADSDNVFITTPDLSLSERLAAQAQVYWMALLARDIPFGEFAKNDYVRLAEN 238
Qy 287 LAGMGPNLDVAVISGDTVPDPSQLFRATFTVGVETGPFVSQLVNSFTIDAITVEPKQ 346
Db 239 LQSLPAFKGLN-IPRSEGGKIDPVTDLFRITTPGVVTPGVVSPQMLSDFLIDSIKVTPKA 297
Qy 347 ETAPDLNVVDPEWLNIONGGPPAGPEELDELRNARDLARVSFVDNINTEAYRG 406
Db 298 DPLTPGVYDTAFQPLWLDVQNGASKLETTTDEENPRFIRNGRDATIALRDQLYTEAFRA 357
Qy 407 SLILLELGAFSRPGINGFIDSDRQAGFVNFGTSHYFRLICAAELAQASCYQKWVHRF 466
Db 358 ALILFTGALG--GEVGPYAEAEQOGFATFGEPHILTAMASASSSTRHAWYAKWQVHRM 415
Qy 467 ARPEALGTTLHNTAGDLDADFDISLENDLELLKRVAEINAAQNPNNEVTVLLPQAIQVG 526
Db 416 LRPEAYGALVHTLMRDVITPLPDSILRNTLLELRNVEVHRNQMNPDCGEKTFLLPMAAQG 475
Qy 527 SPTHSPSPSHATONGAFATVILKALIGLDRCGECFPNPPSPSDGLELINF-----G 579
Db 476 SPTHAPSPSHAINNGAYITALKAFLYGACQKCFPNPVSNDGTRIKYKPSGREIVG 535
Qy 580 AC-----LTYEGEINKLVANVAFGRQMLGIHYRFDGICQGLLIGETITVRLHQEL 629
Db 536 ECYNEKGLVGLTYEGELNKSISANVLLGRSHIGVHRMDGVYALNGETSCVRRLOQEL 595
Qy 630 MTFAE-----BATPFRLFTGEVILKFDQDFTSIDGDMCSGLVTVGVADC 674
Db 596 PGLPEAREVEGKKRRGIDIPATYKFRLYSGKILEYGRNLYKLDGKLCG-AFTGDDDFC 653
```

## RESULT 4

```
ID Q8LLW7 PRELIMINARY; PRT; 598 AA.
AC Q8LLW7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Vanadium-dependent bromoperoxidase.
OS Corallina officinalis (Coral seaweed).
OC Eukaryota; Rhodophyta; Florideophyceae; Corallinales; Corallinaceae;
OC Corallinoideae; Corallina.
OX NCBI_TaxID=35170;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22117781; PubMed=12121762;
RA Carter J.N., Beatty K.E., Simpson M.T., Butler A.;
RT "Reactivity of recombinant and mutant vanadium bromoperoxidase from
the red alga Corallina officinalis.";
RL J. Inorg. Biochem. 91:59-69(2002).
DR EMBL; AF218810; ANA46061.1; -.
DR PDB; 1QHB; 05-JUL-00.
DR GO; GO:0004601; F:peroxidase activity; IEA.
DR InterPro; IPR008934; AcPase_VanPerase.
DR InterPro; IPR000326; PA_PTPase.
KW Peroxidase.
SQ SEQUENCE 598 AA; 65458 MW; BC784E370D748F01 CRC64;
```

Query Match 17.7%; Score 626; DB 10; Length 598;  
Best Local Similarity 31.4%; Pred. No. 1.4e-38;  
Matches 176; Conservative 76; Mismatches 220; Indels 88; Gaps 15;

Qy 168 NGETDLATMFHKSUPHDELGO-VTADDFAILEDCLNGDFSICBDV---PAGDPAGRL-- 221

```
Db 50 NGDPSFTVSTKGLPHDDNGAIIIDPDDFLAFVRAINSGDEKEIADLTGLGPARDPDTGLPI 109
Qy 222 -----VNPTAAFAIDISGPAFSATTIPVPTLSSPELAQAELLYMALARDPVFMQYGTDEITTTAAAN 286
Db 110 WRSDLANSLLEVGRWNSAGLTFLDEGDAQSIAMPAPVLTSPELIAEIAELYLWAL 169
Qy 266 ARDVPF-----MOYGTDEITTTAAANLAGMGFPNLDVAVISGDTVPDPSQLP 314
Db 170 GREIFSEFSPKNAEYIQPAIQNLGLEWFTNPMGLDPPAEIRRRRGVTV---GNLNF 226
Qy 315 RATFVGVTGCTPFVSQLLV-----NSFTIDAITVEPKQET 348
Db 227 RGLPQSEVGPYLSQVYIIVGSKQISATGKNKTLVSPNAADEFGEIAYGSITISQVRRI 286
Qy 349 FAPDLNVVDPEWLNIONGGPPAGPEELDELRNARDLARVSFVDNINTEAYRGSL 408
Db 287 ATGRDFMTDLKVLVDQDAADRGESYEPGARLTIIRDLATVWHFDALYEAYLNACL 346
Qy 409 ILLELGAFSRPGINGFIDSDR---QAGFVNFGTSHYFRLI-GAAELAQASCYQKWQVH 464
Db 347 ILLANRVFPDPNI--PFQEDKLDNDQVFNFGDAHVLSLVTEVATRALKAVRYQKENIH 404
Qy 465 RFARPEALGTTLH-NTIAGD---LDADFDISLENDLELLKRVAEINAAQN-----PNNE 514
Db 405 RRLRPEATGGLISVKNIAEKGSEVPEVDLAVELEULEKAEISNRKQNIADGDDPDD 464
Qy 515 VTLLPQAIQVGSPTHSPSHATONGAFATVILKALIG---LDRGECFPNPPSPSD 570
Db 465 PSFELLQAPAEAGSFPHSPSYGSHAVAGACVTLKAFDFSDNFQIDQVFE-----VDKDED 519
Qy 571 GLELINPEGACLTVEGEINKLVANVAFGRQMLGIHYRFDGICQGLLIGETITVRLHQELM 630
Db 520 KLVKSFKGT-LTAVAGELNKLADNIAIGRNAGVHYFSDQFESILLGEQVAILERQSL 578
Qy 631 TPAREATFPERLFTGEVILK 650
Db 579 TYGENFFFNLPKPDGTTIQI 598
```

## RESULT 5

```
ID O81959 PRELIMINARY; PRT; 598 AA.
AC O81959;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Vanadium-dependent bromoperoxidase 1.
OS Corallina pilulifera.
OC Eukaryota; Rhodophyta; Florideophyceae; Corallinales; Corallinaceae;
OC Corallinoideae; Corallina.
OX NCBI_TaxID=78447;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98307393; PubMed=9645486;
RA Shimomishi M., Kuwamoto S., Inoue H., Wiewer R., Ohshiro T., Izumi Y.,
Tanabe T.;
RT "Cloning and expression of the gene for a vanadium-dependent
bromoperoxidase from a marine macro-alga, Corallina pilulifera.";
RL FEBS Lett. 428:105-110(1998).
DR EMBL; D87657; BAA31261.1; -.
DR HSSP; P81701; 1Q19.
DR GO; GO:0004601; F:peroxidase activity; IEA.
DR InterPro; IPR008934; AcPase_VanPerase.
DR InterPro; IPR000326; PA_PTPase.
KW Peroxidase.
SQ SEQUENCE 598 AA; 65313 MW; 88B81B678B7FACE0 CRC64;
```

Query Match 17.7%; Score 625; DB 10; Length 598;  
Best Local Similarity 31.9%; Pred. No. 1.6e-38;  
Matches 191; Conservative 75; Mismatches 233; Indels 100; Gaps 20;

Qy 141 RDNVAFASRRDAARRERDGTGTVCQITNGE-----TD--LATMFHKSUPHDELG 187

```
Db 11 RAKASFDRVAALALN-ARGVPSFANGELLRYNPDNDTSPSTASTKGLPHDDNG 69
Qy 188 Q-VTADDFAILDCILNGDFSICEDV---PAGDPAGRL-----VNPT 225
Db 70 AIIDPDDFLAFVRAINSQDEKEIADLTGLPARDPETGLPIWRSLANSLELEVRGWNSS 129
Qy 226 AAFADISGPAFSATTIPPVPTLSSPELAQAOLAEYLMALARDVPF-----MOY 274
Db 130 AGLTFDEGGDDAOIAMPPAPVLTSPELVAEIAELMALGRETFEFSEFSPKNAEYIQF 189
Qy 275 GTDEBITTTAAANLACMGGFNNLDVSGSDGTVPFSQLFRATPGVGVETGPFVSQLL--- 331
Db 190 AIDOLNGLEWNTPAKLGDPAPAEIRRRRGVTV---GNLFRGILPGSEVGPYLSQYIIVG 246
Qy 332 ---VNSFTI-----DAITVEPKOETFAPDLNMYVDFDEWL 363
Db 247 SKQIGSATVGNKTLVSPNADEFDGEIAYGSITISQVRVRIATPGRDFMTDLKVFLVQDA 306
Qy 369 GPPAGPEELDEELRFIRNARDLARVSFVDNINTEAYRGSLLLELGAFAFSRPGING 428
Db 307 ADFRGFSYEPGARLIITIRDLATVWHFDALYEAYLNACLIILLANGVPFDP--NLPPQOE 364
Qy 429 DR---QAGVFNFTGSHYFRLI-GAAELAQRASCYQKWQVHRFARPEALGGLTLH-NTIA-- 481
Db 365 DKLDNDQVFNFGSAHVLSLVTEVATRALKAVRYQKFNHRRLEPEATGGLISVKNLAPQ 424
Qy 482 --GDLADADFDLSLENDELKRVAEINAAQN-----PNNEVTYLLPOAIQVGSPTHSYP 534
Db 425 KGESIFPEVLDLAEVDELGILEKAEISNRKQNIADGDDPDPFSLFLLPMAFAEGSPFHPYSY 484
Qy 535 SGHATQNGAFATVLKALIGLDGRGCEFPNVPFSD---DGLLEINRFGACITVEGINKL 591
Db 485 SGHAAVAGACVTLKAF--FDSGLIEI--DQVEFVDEKEDKLKVSFKGT-LTVAGELNKL 539
Qy 592 AVNVAFGRQMLGHIHYRFDGIGLLGETITVTLHQELMTFAEATPEFRLFTGEVILK 650
Db 540 ADNIAIGRNAGVHYFSDQFESLLLGQVAIGILEQSLTYGENFFNLPKFDGTTIQI 598

RESULT 6
O81960 PRELIMINARY; PRT; 597 AA.
AC O81960;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Vanadium-dependent bromoperoxidase 2.
OS Corallina pilulifera.
OC Eukaryota; Rhodophyta; Florideophyceae; Corallinales; Corallinaceae;
OC Corallinoideae; Corallina.
OX NCBI_TaxID=78447;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98307393; PubMed=9645486;
RA Shimonishi M., Kuwamoto S., Inoue H., Wiewer R., Ohshiro T., Izumi Y.,
RA Tanabe T.;
RT "Cloning and expression of the gene for a vanadium-dependent
RT bromoperoxidase from a marine macro-alga, Corallina pilulifera.";
RL FEBS Lett. 428:105-110(1998).
DR EMBL; D87658; BAA31262.1; -.
DR HSP; P81701; 1019.
DR GO; GO:0004601; F:peroxidase activity; IEA.
DR InterPro; IPR008934; ACPase_Vanperase.
DR InterPro; IPR000326; PA_Prtase.
KW Peroxidase.
SQ SEQUENCE 597 AA; 65199 MW; EAD931497060D3EE CRC64;

Query Match 17.4%; Score 614.5; DB 10; Length 597;
Best Local Similarity 31.6%; Pred. No. 1e-37;
Matches 193; Conservative 70; Mismatches 223; Indels 125; Gaps 21;

Qy 141 RDNVAFASRRDARRRDGTGTCVQITNGE-----TDLATMFHKSPLHDELGQ- 188
```

```
Db 11 RAKASFDRVAALAL-ARGVPSFANGELLRYNCETGDPSPFIASFTKGLPHDDNGAI 69
Qy 189 VTADDFAILDCILNGDFSICEDV---PAGDPAGRL-----VNPTAA 227
Db 70 IDPDDFLAFVRAINSQDEKEIADLTGLPARDPETGLPIWRSLANSLELEVRGWNSSAG 129
Qy 228 AAFADISGPAFSATTIPPVPTLSSPELAQAOLAEYLMALARDVPFQYGTDE---ITTTAA 284
Db 130 LTDLSEGDDAQSVAMPAPVLMSPELIAEAEYLYMALGRDIEFSEFSPKNAAFIRSAI 189
Qy 285 ANLAGMGGFNNLDA-----VSGSDGTVPFSQLFRATPGVGVETGPFVSQL 329
Db 190 ERLNGLEWF-NTPAKLGDPAPAEIRRRRGVTVGN-----LFRGILPGSEVGPYLSQ 239
Qy 330 LL-----VNSFTI-----DAITVEPKOETFAPDLNMYVDFDEWL 363
Db 240 YIIVSKQIGSATVGNKTLVSPNADEFDGEIAYGSITISQVRVRIATPGRDFMTDLKVFL 299
Qy 364 NIQNGGPPAGPEELDEELRFIRNARDLARVSFVDNINTEAYRGSLLLELGAFAFSRPGING 423
Db 300 DVQGDADFRGFSYEPGARLIITIRDLATVWHFDALYEAYLNACLIILLANGVPFDP--NL 357
Qy 424 PFIDSDRQAG---FVNFGTSHYFRLI-GAAELAQRASCYQKWQVHRFARPEALGGLTL--- 476
Db 358 PFQEDKLDNDHDFVNFVNGSAHVLSLVTEVATRALKAVRYQKFNHRRLEPEATGGLISVN 417
Qy 477 -HNTIAGDLADFDLSLENDELKLR-----VAEINAAQNNEV-----TYLLPOAI 523
Db 418 KKSFLAGS-----DIIFPEVSELVEELSSILDVAESNEKQNRADGIVSPDKSFLLPMAF 472
Qy 524 QVGSPTHSYPSGHATQNGAFATVLKALIG---LDRGCEFPNVPFSDGLELINFEG 579
Db 473 AEGSPHPSYSGSHAVVAGACVTLKAFDANFQIDKVE-----VDTDEKLKVSFKG 527
Qy 580 ACITVEGEINKLVANVAFGRQMLGHIHYRFDGIGLLGETITVTLHQELMTFAEATPE 639
Db 528 T-LTVAGELNKLADNVAIGRNAGVHYFSDQFESLLLGQVAIGILEQSLTYGENFFN 586
Qy 640 FRLFTGEVILK 650
Db 587 LPRFEDGTTIQI 597

RESULT 7
Q7UDU8 PRELIMINARY; PRT; 8173 AA.
AC Q7UDU8;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Probable aggregation factor core protein MAFP3, isoform C.
OS Rhodospirillum rubrum.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=1;
RX MEDLINE=22735913; PubMed=12835416;
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA Schlesner H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
RT strain 1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
DR EMBL; BX294153; CAD79309.1; -.
KW Complete proteome.
SQ SEQUENCE 8173 AA; 826160 MW; B7DD4ADC483AC614 CRC64;

Query Match 3.8%; Score 133.5; DB 16; Length 8173;
Best Local Similarity 21.9%; Pred. No. 19;
```



```
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AE005097; AAC20220.1; -.
DR PIR; H84356; H84356.
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR00873; AMP-bind.
DR Pfam; PF00501; AMP-binding; 1.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00455; AMP_BINDING; 1.
KW Ligase; Complete proteome.
SQ SEQUENCE 650 AA; 70748 MW; 2AF0B70543983853 CRC64;

Query Match 3.6%; Score 127.5; DB 17; Length 650;
Best Local Similarity 22.6%; Pred. No. 0.92; Indels 109; Gaps 22;
Matches 97; Conservative 48; Mismatches 176;

QY 10 RGSMPDPTGVLRLITSEORAKGW-----RRQLEGEKSLGFHP-SETPYIKYLEGSETWK 62
DB 237 RFGPRDRGAPAVTSDSRHLSFLPLAHVLERLSGH-----FLPLASGAQVCYAESPTLR 292
QY 63 K---VKLPTDGI SASKILGKIMARVRIATATLAVVLAAPCLAFDEVTVASGVFPPEKHGTGE 119
DB 293 EDFGLPEPTSATSPRVYEKLYDTIREQAA-----DSPVSERVF-EWATRVGR 339
QY 120 GRHLQCTNSDDALDPTAPNR-RDNVAFASRRDAARRD-----GTCTVQIINGETDLA 174
DB 340 AAH-----TTDDPGIGLRAAHALADTLVFSIDRALGGEIEFFISGGSL-----SPELC 389
QY 175 TMFH-KSLPHDELQGVTTADDFAILEDCLNGDPSICEDVPAGDPAGRLVNPNTAAFAIDIS 233
DB 390 ALFHGMLP-----LLEG-YLLETSPV-----LAVNPYEDPIVGTI 425
QY 234 GPAPSAITIPVPPLTSPE-----LAAQLAELYWALARDVPFMQYGTDEI 279
DB 426 GPPVTDTLTVDETIASPEQRQRCGGAAGELLARGPQVFDGYW-----GLPDA 473
QY 280 TTTAAANLAGMGFPNLDVSGDGTV---DPSQLFR-ATFVGVETGPFVSQLVNSF 335
DB 474 TDAFVTRGKEWFRGTGDVVELRPDGYRFLERAKQLLTSTGNKVPAGPIEDAFASPL 533
QY 336 TIDAITVEPKQETFAPDLNTVMVDFD---EMLNITQNGGPPAGPELOBELAFIRNARDLAR 392
DB 534 VAQMVVGDQKQFVSAIL--VPNFDAVSEWAASQEIATP-----DDRDAICRDERVAR 585
QY 393 V-SFVDNINT 401
DB 586 IQSAVDDVNT 595

RESULT 10
ID Q89UJ13 PRELIMINARY; PRT; 4210 AA.
AC Q89UJ13;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE B115471 protein.
GN B115471.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Igesawa K., Iriiguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpou S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
```

```
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005955; BAC50736.1; -.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR003644; Calx beta.
DR InterPro; IPR000169; SHPOT acsite.
DR Pfam; PF03160; Calx-beta; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
KW Complete proteome.
SQ SEQUENCE 4210 AA; 428715 MW; D6D3B6879B95BAAF CRC64;

Query Match 3.6%; Score 125.5; DB 16; Length 4210;
Best Local Similarity 20.6%; Pred. No. 27;
Matches 166; Conservative 93; Mismatches 273; Indels 275; Gaps 45;

QY 6 ADTTRGSPMPD-TGVRLRLTSEORAKWRRLQLEGEKSLGHPSE--TPYIKYLEGSETWK 62
DB 1080 ADETQGNQANDVTGPIQIFA-----GVQNPQGVDPDEPGTP-LAYATSSGAAL 1125
QY 63 KVKLPDGTGISASKILGKIMARVRI-----ATLAVVLAAPCLAFDEVTVASGVFP 111
DB 1126 SI-LPYFGVDGPDPAAAIITYSLEKSSNFGVTLTDGTAISLSDSAGRVITGVTGDNVHP 1184
QY 112 EEKHHTGEGRHLQCTNSDDALDP-----TAPNRDNVAFASRRDAARR 135
DB 1185 E-----LSGTTAFAVAIDPLSGKVSVEYLSLHSDSNPNDOVSLAAGSISA-- 1232
QY 156 ERDGTGTVQIINGETDLATM-----FKSLPHDELQGVTTADDFAILEDCLNGD 205
DB 1233 -----VVTIKDGDGTASADISKIHFNDDGDPALITGTGVTPELVAL---ALNLD 1282
QY 206 FSICED-----VPAGDPAGRLVNPNTAAFAIDISGPAPSAITIPVPPLTSPEL---A 254
DB 1283 ETTHGDSPPDQVNLDESGEGGVPNGTA---DDTGNATVTTTDTATSTAQIGELMTGA 1338
QY 255 AQLAELYWALARDVPFMQYGTD-----EITTTAAANLAG---MGGF-PNLDAVSTGSD 304
DB 1339 GLLAPLFTNAV-----IDYGADGAPALVNPVTSALSILVTGAPEAGGIKTNLTATL-AD 1391
QY 305 GTVDPFSQLFRATFVGVETGPFVSG-----LLVNSFTIDAITVEPKQETFA 350
DB 1392 SLLEGTSGEHRITVWLTIEGQIVGRVGHDTLSTADDFFVLRILSNSTDPATQLVVDQFL 1451
QY 351 PDLNVMVDDEWLNIQNGGPPAG-----PEELDEELRFR----- 385
DB 1452 P-----IDHD-----ASEPAGAQLPENPSLFDSEISLTLTATAGOSVGKLVTVTVDGD 1499
QY 386 --NARDLARVSFVDNINT-----EAYRG-----SLILLELGAFSRPGINGPFI 426
DB 1500 GDHVTKSATATVIDHSNISVSDSGPTTISASQDCVDTPLSYVGIKAGNVDERGLDGP-S 1558
QY 427 DSDRQAGFVNGFTSHYFRLLIG--AELAQRASCYQKWQVHRFARPBALGGTLNHTIAGD- 483
DB 1559 DQDVLMSGVGQGNPNANTINTTSGSDIGVGNQSGIEGHEVH-----GG---NDLAAEI 1607
QY 484 LDADF--DISLLENDELLKRVAEINAAQ-----NPNNEVTYLLPQALQVGSPTHPSY 533
DB 1608 LRLDFVNNVTFPANTFAYNGHYEVDAASTFIHOIGNPSNTATVF---VQV----- 1655
QY 534 PSGHATQNGAFATVLKALIGDRGGECFPNPV-----FPSDDGLELINF- 577
DB 1656 --FNANDNDF-----DDGN---PLPITLNDVTVTGDASYTKIPYVDGNTLVGVV 1701
QY 578 -----EGACTYB--GEINKLAV-----NVAFG-----RQMLGIHYRFR--D 609
DB 1702 IVGLNEGATVTVDTSNDFNRLVISNDVAFSTPNNGAPTLPVDPNDSDQLNMGHPFSVS 1761
QY 610 GIGILLGETITVRLHQLMTPAEAA 636
DB 1762 GISSVVCAFA-TLSVTHDESAGAPQS 1787
```





```
QY 490 ISLENDLL-KRVAEINAAQNPNEVYLLPQAIQVGSPTHSYP-----SGH 537
Db 757 VNLNRCGLQKNVIAQSDDTPTGFCYLAESINDKGRSNFSQPVGVGEIRIKLDGE 816
QY 538 ---ATQNGAFATVVKALIGLDRGCECFPNPV 565
Db 817 YQPLKDNVSVTVSKD-SGIVKGAESFTAPV 846

RESULT 13
Q9L8C6
ID Q9L8C6 PRELIMINARY; PRT; 3798 AA.
AC Q9L8C6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Polyketide synthase.
GN EPD.
OS Polyangium cellulosum.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Sorangineae; Polyangiaceae; Polyangium.
OX NCBI_TaxID=56;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=So ce90;
RX MEDLINE=20130945; PubMed=10662695;
RA Molnar I., Schupp T., Ono M., Zirkle R.E., Milnanow M., Cyr D.D.,
RA Nowak-Thompson B., Engel J.M., Toupke C., Stratmann A., Ligon J.M.,
RA Gorlach J., Mayo J.M., Hu A., Goff S., Schmid J., Ligon J.M.;
RT "The biosynthetic gene cluster for the microtubule-stabilizing agents
RT epothilones A and B from Sorangium cellulosum So ce90.";
RL Chem. Biol. 7:97-109(2000).
DR EMBL; AF210843; AA26922.1; -.
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. . .; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001227; Ac trans.
DR InterPro; IPR000794; Ketoacyl synth.
DR InterPro; IPR001601; Methyltransf.
DR InterPro; IPR002025; NAD BS.
DR InterPro; IPR006163; Pp Bind.
DR InterPro; IPR000051; SAM bind.
DR Pfam; PF00698; Acyl transf; 2.
DR Pfam; PF00109; ketoacyl-synt; 2.
DR Pfam; PF02801; ketoacyl-synt_C; 2.
DR Pfam; PF00550; pp-binding; 2.
DR PROSITE; PS00075; ACP DOMAIN; 2.
DR PROSITE; PS00606; B_KETOACYL SYNTHASE; 2.
KW Phosphopantetheine, Transferase.
SQ SEQUENCE 3798 AA; 405273 MW; 2033E57B254CC077 CRC64;

Query Match 3.5%; Score 123.5; DB 2; Length 3798;
Best Local Similarity 21.9%; Pred. No. 32;
Matches 109; Conservative 51; Mismatches 185; Indels 153; Gaps 22;

QY 158 DGTGTCQITNGETDLMTHFKSLPHDELQVTDADDPFAILEDG-----200
Db 2760 DGEHFVSQPLPEPDLAVAL-----EAGRVFADLPVLFEWCKFAGERLADVLTKTLA 2813
QY 201 --IL--NGDPSICDVPAGDP-----AGRLVNPTAAFAIDISGPAFSATT 241
Db 2814 LEILPFGSGFDMAEIRYDSDPIARYSGVIGCVESARVAVPSCMSILEIGTGTATT 2873
QY 242 IPPVPTLSSPELAQAQLAYMWALA-----RDVPFMQYGTDEITTTTAAANLAGMGFPN 295
Db 2874 AAVLPVLLPRTVEHFTDVSPLFLARAEQRPRDYPFLKYGLDQVDPAGQVQAHRQFDV 2933
QY 296 LDAVSGSDGTVDPFSQLFRATFVGVTGTFVSVOLLVNSFTIDAITVEPKQETPAP-DIN 354
Db 2934 IVAANV-IHATRD-----IRAT-----AKRL-----SLLAPGGLL 2963
QY 355 YMVD---FDEWLNIQGGPPAGPELDELRF-----IRNARD-----LARVSFVDNI-----399

RESULT 14
Q84HP6
ID Q84HP6 PRELIMINARY; PRT; 1939 AA.
AC Q84HP6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE PKSE.
GN PKSE.
OS Amycolatopsis orientalis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Pseudonocardineae; Pseudonocardaceae; Amycolatopsis.
OX NCBI_TaxID=31958;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 43491;
RX MEDLINE=22447897; PubMed=12536216;
RA Zazopoulos E., Huang K., Staffa A., Liu W., Bachmann B.O., Nonaka K.,
RA Alelet J., Thorson J.S., Shen B., Farnet C.M.;
RT "A genomics-guided approach for discovering and expressing cryptic
RT metabolic pathways.";
RL Nat. Biotechnol. 21:187-190(2003).
DR EMBL; AF546139; AA025836.1; -.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001227; Ac trans.
DR InterPro; IPR000794; Ketoacyl synth.
DR Pfam; PF00698; Acyl transf; 1.
DR Pfam; PF00109; ketoacyl-synt; 1.
DR Pfam; PF02801; ketoacyl-synt_C; 1.
DR PROSITE; PS00606; B_KETOACYL SYNTHASE; 1.
SQ SEQUENCE 1939 AA; 203659 MW; F0BB981562210E09 CRC64;

Query Match 3.5%; Score 123; DB 2; Length 1939;
Best Local Similarity 22.6%; Pred. No. 12;
Matches 146; Conservative 59; Mismatches 249; Indels 192; Gaps 32;

QY 4 HAADTTTSGPMPTGTVLRLTSTQRAK---GWRQLEGEKSLGFHPSETPIYKLGSET 60
Db 606 HADDIYRAANL-STGADQVATDVAQPRIVTGSAGLRVLKSLGTEATV--TGHSLGELT 662
QY 61 ---WKKVKLPDTCISASKILKIMARV-----RIATALA-----91
Db 663 ALHWGALTREVLKUKAKIRGKYNATASDGDGAMAAIAATPSPAEGLAEEVVIAGNA 722
QY 92 ---VLAAPCLAFDEVTA-----SGVFFEEHKHGTGGRHLQ---124
Db 723 PEQTVLSPAEADRVVARARAEVTAARINVSHAFSPAVVPAEAAMTCELAAIDFARL 782
QY 125 ----TCTNSDDALDPTAPNR---RDNVAFASR-RDAARRRDGTGTVCQITNCE-----170
Db 783 DRPVWSTVTGDLVHAAEDLRLDLRDQVLPVRFREAAKVAERSDLVIEVGPGRVLTGLL 842
```

```
QY 171 -----TDLATMFHKSILPHDELQVTTADFAI-----LEDCL-----NGDFSI-- 208
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 843 GTIAPGTPVLSDTSLTAPVLKVGAGAFAGAQLESTLFDGRVVRALPADGEFSFLA 902
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 209 ---CEDVP---AGDPAGRLVNPTAFATIDISGPASATTIPVPTLSPPELAQAELIYWM 263
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 903 SPCEAASISGAVLTRDVAEPAEAAAGTASGSGST-----LDLLR--KLASERVELPLE 956
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 264 ALARDV--PFMYQGTDEIT-----TTAAANLACMGFPNLDAYSIGS-DGTVDPPSQLF 314
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 957 AVTADTHPLDLHLSSITVGLVNDVTRALGRPALEGMFPAFATVCLGELAEMIDELAOTA 1016
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 315 R-----ATFVGVEGPFVSQLVNSFTIDAIVPEKQETFPAPLNVMPVDFEVLNIONG 368
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1017 KPADSNOAEVAGV--GFW-----VRPEAVEV--VAPKP---SPDLATGISTAEWTAPAPA 1065
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 369 GPPAGPPELDEELRFRINARDLARVSVDNINTEAYRGSLLILLELGFAPSPGNGPFIDS 428
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1066 GHP-----LAELPLR-----AALATAGVGDG-----VLLCLNADSASGDVGLFLDA 1105
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 429 DR-----OAGFVNFGTSHYFRL-----IGAAELAQASCYQKQVHR 465
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1106 GRAVLAAPNGTRFVVVOHGLGASLAKTLRLDPSARTTIVDLADLGP----- 1153
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 466 FARPEALGGTLHNTIAGLDADDFDISLENDELKRVAEINAAQNP 511
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1154 -VDPEALDAAV-STVVTVEAATDFSEVRYDTAGVRTVPKLAALTP 1197
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
Q9FCJ6 PRELIMINARY; PRT; 469 AA.
ID AC Q9FCJ6
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE Hypothetical protein SC05192.
GN SC05192 OR 28C3B6.16C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,
```

```
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939122; CAC01322.1; -.
DR InterPro; IPR004147; ABC1.
DR Pfam; PF03109; ABC1; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 469 AA; 51306 MW; A22E1FD8324E260C CRC64;

Query Match 3.5%; Score 122.5; DB 16; Length 469;
Best Local Similarity 20.4%; Pred. No. 1.3;
Matches 76; Conservative 57; Mismatches 121; Indels 119; Gaps 18;

QY 226 AAFADISGPAFSATT-----IPPVPTLSSPELAQAELIYWMALARDVPMOYGTDEIT 280
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 77 SALPEEVAGYRAALTQLQEAAPMPTRTVHSLAERLGEDWQEL-----FLEF--EDK 128
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 281 TTAANL-----ACMGFPNLDAYSIGSDGTVDPFPSQLFRATFVGVTGPF 326
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 129 PAAAASIGQVHRGWODGREVAVKVQYPGAGEALLSDLNQLSRFARLLGLPLVPGMDIKPL 188
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 327 VSOLLVN-----SFTIDAIVPEKQETF-----APDLNMYD---FDEWLN----- 364
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 189 ITELKORVSEELDYDLQEAQAQRTAEVFDPPDIVVDPDVHQCEQVLITEMIDGIPMSEI 248
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 365 IQNGGPPAGPEELDEELRFRINARDLARVSF-----VDNINTEAYRGSLLILLELGFAPSRP 419
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 249 ISNG-----TOEQRD-----RAGQLLAHFLFSGPARTGLLHADPHGPNFRLL-----P 291
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 420 GINGPFIDSDRQAGFVNFGTSHYFRLIGAAELAQASCYQKQVHRFARPEALGGTLHNT 479
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 292 G--GPRGEDDWRLGLVDFGTVD--RLPGGL-----PEPIGEALRMT 328
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 480 IAGDLADDFDISLENDELKRVAEINAAQNPNNEVTYLLP----- 520
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 329 LDGEAEAVYEMLCAG--FVKESIEL-----APDAVLEYLLPIIEPARVDAFTFTRGWRS 382
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 521 QAIQVGSPTHPSPY 533
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 383 QAARIGDPRSPAY 395
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: September 19, 2004, 03:58:47  
Job time : 129.3 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2004, 01:14:52 ; Search time 7.02288 Seconds  
(without alignments)  
1468.042 Million cell updates/sec

Title: US-10-691-383-2\_COPY\_435\_632

Perfect score: 1039

Sequence: 1 VNFQTHYFLIGABLAQR.....GLLGGTTIVTLHQLMLTF 198

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	935	90.0	557	1 PRXV_ASCNO	P81701 ascopyllium
2	85.5	8.2	609	1 PRXC_CURIN	P49033 curvularia
3	85.5	8.2	1819	1 GCP6_HUMAN	Q96rt7 homo sapien
4	82.5	7.9	330	1 Y374_METJA	Q57819 methanococ
5	82	7.9	567	1 RST2_SCHPO	P78871 schizosacch
6	79.5	7.7	198	1 YBJG_ECOLI	P75806 escherichia
7	79.5	7.7	419	1 MUA2_STRP8	Q8p097 streptococc
8	79.5	7.7	419	1 MUA2_STRPY	Q99z78 streptococc
9	79.5	7.7	922	1 YB29_HUMAN	Q9hb77 homo sapien
10	78.5	7.6	206	1 Y236_THEAC	Q9hlj2 thermoplasm
11	77.5	7.5	454	1 SR54_AQUAE	O67615 aquifex aeo
12	77.5	7.5	940	1 PTGA_MYCPN	P75569 mycoplasma
13	77	7.4	495	1 TOLC_ECOLI	P02930 escherichia
14	76	7.3	394	1 PGK_BACME	P24289 bacillus me
15	75.5	7.3	426	1 APEB_MYCLE	Q50022 mycobacteri
16	75.5	7.3	493	1 GALT_LACIL	Q98683 lactococcus
17	75	7.2	337	1 GLK_XYLFT	O87es6 xyliella fas
18	75	7.2	774	1 STF_LAMED	P03764 bacterioph
19	74.5	7.2	260	1 OPRI_NEIMC	P10170 neisseria m
20	74.5	7.2	835	1 VIRL_AGRU1	P07167 agrobacteri
21	74	7.1	237	1 YEIU_ECOLI	P76445 escherichia
22	74	7.1	611	1 GLMS_RALSO	Q8y303 r glucosami
23	73.5	7.1	154	1 FMM_MORNO	Q9829 moraxella n
24	73.5	7.1	472	1 GCSB_THEVO	Q97c04 thermoplasm
25	73.5	7.1	803	1 DCML_HYDPS	P19913 hydrogenoph
26	73	7.0	603	1 LEPA_SINY3	P74751 synochocyst
27	73	7.0	1163	1 Y222_HUMAN	Q92618 homo sapien
28	72.5	7.0	216	1 Y007_TREPA	Q83052 treponema p
29	72.5	7.0	304	1 STKG_MOUSE	Q88697 mus musculu
30	72.5	7.0	316	1 ARCC_BACLI	O86134 bacillus li
31	72.5	7.0	398	1 A23D_DROME	Q24093 drosophila
32	72.5	7.0	417	1 SAHH_METHF	O27673 methanobact
33	72.5	7.0	630	1 GIDA_PSEPU	P25756 pseudomonas

## RESULT 1

ID	PRXV_ASCNO	STANDARD;	PRT;	557 AA.
AC	P81701;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Vanadium haloperoxidase (EC 1.11.1.-) (V-BPO).			
OS	Ascopyllium nodosum (Knotted wrack) (Brown seaweed).			
OC	Eukaryota; stramenopiles; Phaeophyceae; Fucales; Fucaeeae;			
OC	Ascopyllium.			
OX	NCBI_TaxID=52969;			
RN	[1]			
RP	SEQUENCE, X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS), AND FUNCTION.			
RX	MEDLINE=20013071; PubMed=10543953;			
RA	Weyand M., Hecht H.-J., Kiess M., Liaud M.-F., Vilter H., Schomburg D.;			
RA	"X-ray structure determination of a vanadium-dependent			
RT	haloperoxidase from Ascopyllium nodosum at 2.0-A resolution."			
RL	J. Mol. Biol. 293:595-611(1999).			
RN	[2]			
RP	SEQUENCE OF 320-556 FROM N.A., SEQUENCE OF 326-341; 383-426; 471-479			
RP	AND 481-556, AND FUNCTION.			
RX	MEDLINE=96081028; PubMed=8564812;			
RA	Vilter H.;			
RT	"Vanadium-dependent haloperoxidases."			
RL	(In) Sigel H., Sigel A. (eds.);			
RL	Metal ions in biological system-vanadium and its role in life,			
RL	pp.31-325-362, Marcel Dekker, New York (1995).			
CC	-1- CATALYTIC ACTIVITY: Halide + H(2)O(2) + H(+) = HOHal + H(2)O.			
CC	-1- COFACTOR: Vanadium.			
CC	-1- SUBUNIT: Homodimer; disulfide-linked.			
CC	-1- SIMILARITY: TO OTHER BACTERIAL NON-HEME BROMO- AND CHLORO-			
CC	PEROXIDASES.			
DR	PDB; 1O19; 10-JUN-00.			
DR	InterPro; IPR008934; ACPase_VanPerase.			
DR	InterPro; IPR000326; PA_PPase.			
KW	Oxidoreductase; Peroxidase; Vanadium; 3D-structure;			
KW	Pyrrrolidone carboxylic acid.			
FT	MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.			
FT	DISULFID 3 3 INTERCHAIN (WITH C-41).			
FT	DISULFID 41 41 INTERCHAIN (WITH C-3).			
FT	DISULFID 77 86			
FT	DISULFID 441 462			
FT	DISULFID 544 555			
FT	ACT_SITE 411 411			
FT	ACT_SITE 418 418			
FT	METAL 486 486			
FT	CONFLICT 321 321 S -> D (IN REF. 2).			
FT	CONFLICT 341 341 K -> N (IN REF. 2).			
FT	CONFLICT 403 404 AI -> VY (IN REF. 2).			
FT	CONFLICT 407 408 GS -> T (IN REF. 2).			
FT	CONFLICT 409 409 P -> S (IN REF. 2; AA SEQUENCE).			
FT	CONFLICT 441 444 CYPD -> AIR (IN REF. 2).			
FT	CONFLICT 470 470 N -> K (IN REF. 2).			
FT	HELIX 15 37			



```

50 52
53 53
56 57
60 60
62 63
66 66
68 80
83 87
88 88
98 100
102 103
105 108
118 119
129 130
132 146
147 149
152 154
155 157
159 170
172 173
174 176
178 179
180 180
184 185
190 195
197 197
200 203
210 213
216 218
219 220
221 223
227 227
230 230
236 236
240 247
248 249
258 258
263 263
268 277
282 293
294 295
299 300
304 308
312 313
315 315
316 317
318 318
320 329
330 333
334 342
343 345
347 347
350 361
362 363
371 374
375 375
377 390
392 393
402 402
405 406
417 433
435 438
439 439
441 441
446 448
450 451
455 457
464 464
466 466
484 484
488 488
519 522
524 525

FT HELIX 50 52
FT STRAND 53 53
FT TURN 56 57
FT STRAND 60 60
FT TURN 62 63
FT STRAND 66 66
FT TURN 68 80
FT HELIX 83 87
FT TURN 88 88
FT STRAND 98 100
FT TURN 102 103
FT STRAND 105 108
FT TURN 118 119
FT STRAND 129 130
FT TURN 132 146
FT STRAND 147 149
FT TURN 152 154
FT STRAND 155 157
FT TURN 159 170
FT STRAND 172 173
FT TURN 174 176
FT STRAND 178 179
FT TURN 180 180
FT STRAND 184 185
FT TURN 190 195
FT STRAND 197 197
FT TURN 200 203
FT STRAND 210 213
FT TURN 216 218
FT STRAND 219 220
FT TURN 221 223
FT STRAND 227 227
FT TURN 230 230
FT STRAND 236 236
FT TURN 240 247
FT STRAND 248 249
FT TURN 258 258
FT STRAND 263 263
FT TURN 268 277
FT STRAND 282 293
FT TURN 294 295
FT STRAND 299 300
FT TURN 304 308
FT STRAND 312 313
FT TURN 315 315
FT STRAND 316 317
FT TURN 318 318
FT STRAND 320 329
FT TURN 330 333
FT STRAND 334 342
FT TURN 343 345
FT STRAND 347 347
FT TURN 350 361
FT STRAND 362 363
FT TURN 371 374
FT STRAND 375 375
FT TURN 377 390
FT STRAND 392 393
FT TURN 402 402
FT STRAND 405 406
FT TURN 417 433
FT STRAND 435 438
FT TURN 439 439
FT STRAND 441 441
FT TURN 446 448
FT STRAND 450 451
FT TURN 455 457
FT STRAND 464 464
FT TURN 466 466
FT STRAND 484 484
FT TURN 488 488
FT STRAND 519 522
FT TURN 524 525

Query Match 90.0%; Score 935; DB 1; Length 557;
Best Local Similarity 89.4%; Pred. No. 6.9e-81;
Matches 177; Conservative 12; Mismatches 9; Indels 0; Gaps 0;

QY 1 VNFQSHYFRLIGAAELAQACYQKQVHRFARPEALGGLHTNTIAGDLDADFDISLLE 60
DB VNFQSHYFRLIGAAELAQACYQKQVHRFARPEALGGLHTNTIAGDLDADFDISLLE 375
QY 61 NDELLKRVAEINAAQNPNEVTVLLPOAIOGVSPTSPSGHATQNGAFATVVKALIGL 120
DB NDELLKRVAEINAAQNPNEVTVLLPOAIOGVSPTSPSGHATQNGAFATVVKALIGL 435
QY 121 DRGCECPNPVFPSSDDGLINFEAGCLTYEGEINKLVNVAFCROWLGHYRFDGTOGL 180
DB DRGCECPNPVFPSSDDGLINFEAGCLTYEGEINKLVNVAFCROWLGHYRFDGTOGL 495
QY 181 LLGETITVTRLHOELMTFF 198
DB LLGETITVTRLHOELMTFF 513

RESULT 2
PRXC_CURIN STANDARD; PRT; 609 AA.
AC P49053;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vanadium chloroperoxidase (EC 1.11.1.10) (VCPO) (Vanadium chloride
peroxidase).
GN CPO.
OS Curvularia inaequalis.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
OC Pleosporales; Pleosporaceae; Cochliobolus.
OX NCBI_TaxID=38902;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=95262722; PubMed=7744081;
RA Simons B.H., Barnett P., Vollenbroek E.G.M., Dekker H.L.,
RA Muijsers A.O., Messerschmidt A., Wever R.;
RT "Primary structure and characterization of the vanadium
RT chloroperoxidase from the fungus Curvularia inaequalis.";
RL Eur. J. Biochem. 229:566-574(1995).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
RX MEDLINE=96133943; PubMed=8552646;
RA Messerschmidt A., Wever R.;
RT "X-ray structure of a vanadium-containing enzyme: chloroperoxidase
RT from the fungus Curvularia inaequalis";
RL Proc. Natl. Acad. Sci. U.S.A. 93:392-396(1996).
CC -I- CATALYTIC ACTIVITY: 2 RH + 2 chloride + H(2)O(2) = 2 RCl + 2
CC H(2)O.
CC -I- COFACTOR: Vanadium.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- PTM: The N-terminus is blocked.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X85369; CAA59686.1; -
DR
```

DR PIR; S69334; S69334.  
DR PDB; 1VNC; 08-NOV-96.  
DR PDB; 1IDQ; 11-APR-01.  
DR PDB; 1IDQ; 11-APR-01.  
DR PDB; 1VNE; 11-AUG-99.  
DR PDB; 1VNF; 11-AUG-99.  
DR PDB; 1VNG; 11-AUG-99.  
DR PDB; 1VNH; 11-AUG-99.  
DR PDB; 1VNI; 11-AUG-99.  
DR PDB; 1VNS; 11-AUG-99.  
DR InterPro; IPR008934; AcPase\_VanPerase.  
DR InterPro; IPR003320; Cl\_Perox.  
DR InterPro; IPR000326; PA\_PiPase.  
DR Pfam; PF02328; Cl\_perox; 1.  
KW Oxidoreductase; Peroxidase; Vanadium; Chloride; 3D-structure.  
FT ACT SITE 404 404  
FT METAL 496 496  
FT CONFLICT 454 454  
FT HELIX 16 20  
FT HELIX 22 40  
FT TURN 41 41  
FT HELIX 47 68  
FT STRAND 79 79  
FT TURN 80 81  
FT HELIX 85 87  
FT STRAND 88 88  
FT TURN 93 94  
FT HELIX 98 114  
FT HELIX 131 147  
FT TURN 149 150  
FT TURN 154 155  
FT HELIX 157 173  
FT TURN 176 179  
FT TURN 182 183  
FT TURN 190 191  
FT STRAND 192 192  
FT TURN 197 198  
FT STRAND 202 207  
FT TURN 208 209  
FT STRAND 210 218  
FT TURN 224 225  
FT HELIX 226 229  
FT TURN 245 246  
FT HELIX 248 250  
FT TURN 251 252  
FT HELIX 254 268  
FT STRAND 269 269  
FT TURN 271 272  
FT TURN 274 275  
FT STRAND 280 288  
FT TURN 289 289  
FT STRAND 294 294  
FT TURN 295 297  
FT STRAND 298 298  
FT HELIX 300 314  
FT TURN 315 315  
FT HELIX 325 356  
FT STRAND 358 358  
FT HELIX 361 366  
FT TURN 367 367  
FT TURN 371 373  
FT TURN 376 377  
FT STRAND 380 380  
FT STRAND 383 383  
FT TURN 386 387  
FT STRAND 393 393  
FT HELIX 403 419  
FT STRAND 420 420  
FT TURN 421 423  
FT STRAND 424 424  
FT TURN 428 429  
FT STRAND 436 440  
FT HELIX 442 444

FT STRAND 449 451  
FT TURN 457 458  
FT HELIX 461 463  
FT STRAND 466 468  
FT STRAND 473 476  
FT HELIX 479 492  
FT TURN 493 494  
FT HELIX 498 501  
FT STRAND 504 507  
FT STRAND 508 508  
FT STRAND 510 510  
FT TURN 514 515  
FT STRAND 517 517  
FT STRAND 519 519  
FT TURN 521 522  
FT STRAND 525 525  
FT STRAND 527 527  
FT HELIX 530 532  
FT STRAND 539 539  
FT STRAND 548 548  
FT HELIX 552 565  
FT TURN 566 567  
FT HELIX 573 575  
SQ SEQUENCE 609 AA; 67530 MW; A7B710DDF937D3E9 CRC64;  
Query Match 8.2%; Score 85.5; DB 1; Length 609;  
Best Local Similarity 21.7%; Pred. No. 2.7;  
Matches 30; Conservative 22; Mismatches 35; Indels 51; Gaps 5;  
Qy 71 INAAQNNNEVYLLPOAIQVSGTHPSYSGHATQNGAFATVLK-----115  
Db 381 LGAPATNTDIPF-----KPPFPAYPSGHATFGGAVFQMVRRYNGRVGTWKDDEP 431  
Qy 116 -----ALIGLDRGCECPNPVPSPDGLLELI-----NFEAC-LTYGEINKL 157  
Db 432 DNTAIDNMISEELNGVNRDLRQPDYDTAPIEDQGIIVTRIVRHSWELMFENATISRI 491  
Qy 158 AVNVAFGRQMLGHIHYRED 175  
Db 492 -----FLGVHVRFD 500  
RESULT 3  
GCP6 HUMAN STANDARD; PRT; 1819 AA.  
AC Q96RT7; Q9BY91; Q9UGX3; Q9UGX4;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Gamma-tubulin complex component 6 (GCP-6).  
GN TUBGCP6 OR GCP6 OR KIAA1669.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1), AND CHARACTERIZATION.  
RX MEDLINE=21551508; PubMed=11694571;  
RA Murphy S.M., Preble A.M., Patel U.K., O'Connell K.L., Dias D.P.,  
RA Moritz M., Agard D., Stults J.T., Stearns T.;  
FT "GCPs and GCP6: two new members of the human gamma-tubulin complex.";  
RL Mol. Biol. Cell 12:3340-3352(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20057165; PubMed=10591208;  
RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,  
RA Clamp M., Smail L.J., Ainscough R., Almeida J.P., Babbage A.K.,  
RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,  
RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,  
RA Burdill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,  
RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,  
RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,  
RA Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,

Evans K.L., Fey J.M., Fleming K., French L., Garner A.A., Gilbert J.G.R., Goward M.E., Grahame D.V., Griffiths M.N.D., Hall C., Hall R.E., Hall-Tamlyn G., Heathcott R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A., Laird G.K., Langford C.F., Leverisha M.A., Lloyd C., Lloyd D.M., Martyn I.D., Madsen-Mohammadi M., Matthews L.H., Mccann O.T., Mcclay J., McLaren S., McMurray A., Milne S.A., Mortimore B.J.C.T., Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T., Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T., Scott C.E., Sehra H.K., Stuce C.D., Smalley S., Smith M.L., Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M., Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L., Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L., Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N., Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S., Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T., Dorman A., Fang F., Hu Y., Hua A., Kenton S., Lai H., Lao H.I., Lewis J., Lewis S., Lin S.-P., Loh P., Mala J.E., Nguyen T., Pan H., Phan S., Qi S., Qian Y., Ray L., Ren O., Shaull S., Sloan D., Song L., Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z., Zhan M., Zhang G., Chisole S., Murray J., Miller N., Minx P., Fulten R., Johnson D., Benis G., Bentley D., Bradshaw H., Bourne S., Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozerky P., Rohlfing T., Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J., Korf I., Bedell J.A., Hillier L.W., Mardis E., Waterston R., Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P., Peyrard K., Kedar D., Seroussi E., Franconi I., Tapia I., Bruder C.E., O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L., Tilahun Y., Wright H.;  
 "The DNA sequence of human chromosome 22.";  
 Nature 402:489-495(1999).  
 [3]  
 SEQUENCE OF 33-1819 FROM N.A. (ISOFORM 2).  
 TISSUE=Brain;  
 MEDLINE=21156230; PubMed=11258795;  
 RA Hirosewa M., Nagase T., Murahashi Y., Kikuno R., Ohara O.;  
 RT "Identification of novel transcribed sequences on human chromosome 22  
 by expressed sequence tag mapping.";  
 RL DNA Res. 8:1-9(2001).  
 CC -1- FUNCTION: Gamma-tubulin complex is necessary for microtubule  
 nucleation at the centrosome  
 CC -1- SUBUNIT: Gamma-tubulin complex is composed of gamma-tubulin, GCP2,  
 GCP3, GCP4, GCP5 and GCP6.  
 CC -1- SUBCELLULAR LOCATION: Centrosome.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=Q96RT7-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q96RT7-2; Sequence=VSP\_001624;  
 CC Note=No experimental confirmation available;  
 CC -1- SIMILARITY: Belongs to the GCP family.  
 CC -1- CAUTION: Ref.3 sequence differs from that shown due to frameshifts  
 in positions 1371 and 1758.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 -----  
 CC EMBL; AF272887; AAK82968.1; .  
 CC EMBL; AL022328; CAB63046.1; ALT\_SEQ.  
 CC EMBL; AL022328; CAB63047.1; ALT\_SEQ.  
 CC EMBL; AB051456; BAB33339.1; ALT\_FRAME.  
 CC Genbank; HGNC:18127; TUBGCP6.  
 CC GO; GO:0008274; C:gamma-tubulin ring complex; IDA.

DR GO; GO:0008017; F:microtubule binding; IDA.  
 DR GO; GO:0007020; P:microtubule nucleation; IDA.  
 DR InterPro; IPR007259; SPC97\_Spc98.  
 DR Pfam; PF041130; SPC97\_Spc98; 1.  
 KW Microtubule; Repeat; Alternative splicing.  
 FT DOMAIN 1027 1269 9 x 27 AA TANDEM REPEATS.  
 FT REPEAT 1027 1053 1.  
 FT REPEAT 1054 1080 2.  
 FT REPEAT 1081 1107 3.  
 FT REPEAT 1108 1134 4.  
 FT REPEAT 1135 1161 5.  
 FT REPEAT 1162 1188 6.  
 FT REPEAT 1189 1215 7.  
 FT REPEAT 1216 1242 8.  
 FT REPEAT 1243 1269 9.  
 FT VARSPLIC 1724 1757 Missing (in isoform 2).  
 FT /FTID=VSP\_001624.  
 FT CONFLICT 567 567 S -> L (IN REF. 3).  
 FT CONFLICT 1377 1377 A -> T (IN REF. 2 AND 3).  
 FT CONFLICT 1621 1621 L -> V (IN REF. 2).  
 SQ SEQUENCE 1819 AA; 200455 MW; 923576544D34594A CRC64;  
 Query Match 8.2%; Score 85.5; DB 1; Length 1819;  
 Best Local Similarity 28.9%; Pred. No. 11;  
 Matches 43; Conservative 18; Mismatches 55; Indels 33; Gaps 9;  
 QY 7 HYFLRICAEALASQYOK--WQVHRFARPEALGGTLHTNTIAGDLDDADFDISLEND-- 62  
 DB 763 HYSLK--SAAARRE---QKALMRIOR-----HLESLARL-----FLEDEKH 801  
 QY 63 --ELLKRVAEINAAQNPNNVTVLLPQAIQVGSPTSPSHATQNGAFATVTKALIGL 120  
 DB 802 IQEMLKRVSEAHQPPDPVLLSVHPQVTPSPG-PEHPE--GGQGCDSGSAEQHSAPWDGW 858  
 QY 121 DRGGECPNVPFP---SDDGLLEINFEQA 146  
 DB 859 NRPGLLTTPQKLPLAVGAGRGGLQQAEGA 887  
 RESULT 4  
 Y374\_METJA STANDARD; PRT; 330 AA.  
 AC Q57819;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein MJ0374.  
 GN MJ0374.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcales; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomo J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
 RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii.";  
 RL Science 273:1058-1073(1996).  
 CC -1- SIMILARITY: TO M.JANNASCHII MJ0201 AND TO H.INFLUENZAE AND  
 H.SONNUS HT0703.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its







124 DLHLKAFEAMGVEVSYEGENMNLSTN---GOKIHGAHIYMDTVS---VGATI 169

```

RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Shead M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358(1998).
CC -I- FUNCTION: NECESSARY FOR EFFICIENT EXPORT OF EXTRA-CYTOPLASMIC
CC PROTEINS. BINDS TO THE SIGNAL SEQUENCE WHEN IT EMERGES FROM THE
CC RIBOSOMES (BY SIMILARITY).
CC -I- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 4.5 RNA
CC MOLECULE AND PROTEIN PFH (BY SIMILARITY).
CC -I- DOMAIN: THE PROTEIN HAS A TWO DOMAIN STRUCTURE: THE G-DOMAIN
CC BINDS GTP; THE M-DOMAIN BINDS THE RNA AND ALSO BINDS THE SIGNAL
CC SEQUENCE.
CC -I- SIMILARITY: Belongs to the GTP-binding SRP family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AE000753; AAC07579.1; -.
CC PIR; E70448; E70448.
CC HSP; O07347; IFFH.
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR000897; SRP54.
CC InterPro; IPR004125; SRP54_SPB.
CC InterPro; IPR004780; SRP_sub.
CC Pfam; PF00448; SRP54; 1.
CC Pfam; PF02881; SRP54_N; 1.
CC Pfam; PF02978; SRP_SPB; 1.
CC ProDom; PD000819; SRP54; 1.
CC SMART; SM00382; AAA; 1.
CC TIGRFAMs; TIGR00959; ffh; 1.
CC PROSITE; PS00300; SRP54; 1.
CC Signal recognition particle; GTP-binding; RNA-binding;
CC Complete proteome.
CC
CC DOMAIN 1 289 G-DOMAIN.
CC M-DOMAIN.
CC NP_BIND 102 109 GTP (BY SIMILARITY).
CC NP_BIND 184 188 GTP (BY SIMILARITY).
CC NP_BIND 242 245 GTP (BY SIMILARITY).
CC SEQUENCE 454 AA; 50911 MW; 9f13f8c883a0558 CRC64;
CC
Query Match 7.5%; Score 77.5; DB 1; Length 454;
Best Local Similarity 23.9%; Pred. No. 11;
Matches 43; Conservative 26; Mismatches 40; Indels 71; Gaps 10;
QY 12 IGAAELAQKSCYQKQVHRFAPEALGGLTHTIAGDLDADFDISLENDLLKRVARI 71
Db LGAVEIARRA-----VKR-AKEESVDYLLDT-AGRLHVD-----EELMKELQEI 202
QY 72 NAAQNPNEVYLL-----PQATQVGSPTSPSPSGHATQNGAFATVLLKALIGLDRGEC 126
Db KEVTNP-SEILYVADAMQQTALKTFT-----ERLGTGVVITKQDGDARG-- 251
QY 127 FPNPVPFDDGLELINFEGACLYTEGINKLVAVNFAFGROMLGIHYRFDGIGLIGETI 186
Db -----LALSV---KEVLGPPIKFIGV-----GEKI 273

RESULT 12
PTGA_MYCPN
ID PTGA_MYCPN STANDARD; PRT; 940 AA.
AC P75559;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)

RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Shead M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358(1998).
CC -I- FUNCTION: NECESSARY FOR EFFICIENT EXPORT OF EXTRA-CYTOPLASMIC
CC PROTEINS. BINDS TO THE SIGNAL SEQUENCE WHEN IT EMERGES FROM THE
CC RIBOSOMES (BY SIMILARITY).
CC -I- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 4.5 RNA
CC MOLECULE AND PROTEIN PFH (BY SIMILARITY).
CC -I- DOMAIN: THE PROTEIN HAS A TWO DOMAIN STRUCTURE: THE G-DOMAIN
CC BINDS GTP; THE M-DOMAIN BINDS THE RNA AND ALSO BINDS THE SIGNAL
CC SEQUENCE.
CC -I- SIMILARITY: Belongs to the GTP-binding SRP family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AL445663; CAC11381.1; -.
CC HAMAP; MF 00645; -.
CC InterPro; IPR002733; DUF51.
CC Pfam; PF01871; AMMECR1; 1.
CC ProDom; PD009671; DUF51; 1.
CC TIGRFAMs; TIGR00296; TIGR00296; 1.
CC Hypothetical protein; Complete proteome.
CC SEQUENCE 206 AA; 22971 MW; 5E83B24049A40278 CRC64;
CC
Query Match 7.6%; Score 78.5; DB 1; Length 206;
Best Local Similarity 25.1%; Pred. No. 3.1;
Matches 45; Conservative 19; Mismatches 54; Indels 61; Gaps 10;
QY 52 ADFDISLENDLLKRVAEINAAQNPNNEVYLLPQAIQVGSPTSPSPSGHATQNGAFA 111
Db AEVDISLDIGTKAV-RLRAAASLYRNE---KLDP-----PPADPIFQEKH---GVFT 55
QY 112 TV-----LKALIGLDRGECPPNPVFPDDG-----LELIN 142
Db TINTYPTNLTGCGIG-----PPEYYPILGEGIRSSIVAAEDPPRFPKIDEISHVT 108
QY 143 FEGACLYTEGINKLVN-----VAFGRQMLGIHYRFDGIGLIGETITVRTLHOE 194
Db FEVSILTQPVETI---TWNEDPRKAVHIGRD--GLIAYNGASGLLLPQVATEYRMNPE 162

RESULT 11
SR54_AQUAE
ID SR54_AQUAE STANDARD; PRT; 454 AA.
AC O67615;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Signal recognition particle protein (Fifty-four homolog).
GN FFH OR AQ.1720.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
```

PTS system, glucose-specific IIABC component (EIIABC-GLC) (Glucose-permease IIABC component) (Phosphotransferase enzyme II, ABC component) (EC 2.7.1.69) (EII-GLC/EIIB-GLC).

QY MGSG OR MPN207 OR MP624.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

OX NCBI\_TaxID=2104;

FN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 29342 / M129;

RA MEDLINE=97105885; PubMed=948633;

RX Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C., Hermann R.;

RA "Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae";

RT Nucleic Acids Res. 24:4420-4449 (1996).

RL -!- FUNCTION: This is a component of the phosphoenolpyruvate-dependent sugar phosphotransferase system (PTS), a major carbohydrate active transport system. The IICD domains contain the sugar binding site and the transmembrane channel; the IIA domain contains the primary phosphorylation site (the donor is phospho-HPr); IIA transfers its phosphoryl group to the IIB domain which finally transfers it to the sugar.

CC -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein histidine + sugar phosphate.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein.

CC -!- SIMILARITY: Contains 1 PTS EIIA domain.

CC -!- SIMILARITY: Contains 1 PTS EIIB domain.

CC -!- SIMILARITY: Contains 1 PTS EIIC domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AE000060; AAB96272.1; -.

DR PIR; S73950; S73950.

DR HSP; P05053; 11BA.

DR InterPro; IPR001996; Pfam EIIA.

DR InterPro; IPR003352; Pfam EIIA.

DR InterPro; IPR011127; Pfam EIIA.

DR Pfam; PF00358; Pfam EIIA 1; 1.

DR Pfam; PF00367; Pfam EIIA 1; 1.

DR Pfam; PF02378; Pfam EIIA 1; 1.

DR ProDom; PD002243; Pfam EIIA 1; 1.

DR ProDom; PD001476; Pfam EIIA 1; 1.

DR TIGRFAMs; TIGR00826; EIIA\_glc; 1.

DR TIGRFAMs; TIGR00830; PTBA; 1.

DR PROSITE; PS00371; Pfam EIIA 1; 1.

DR Phosphotransferase system; Sugar transport; Transferase; Phosphorylation; Transmembrane; Complete proteome.

KW DOMAIN 1 7290

FT DOMAIN 1 7290

FT DOMAIN 7291 7446

FT DOMAIN 7447 7628

FT DOMAIN 7629 7730

FT DOMAIN 7731 940

FT TRANSMEM 43 63

FT TRANSMEM 83 103

FT TRANSMEM 112 132

FT TRANSMEM 175 195

FT TRANSMEM 209 229

FT TRANSMEM 238 258

FT TRANSMEM 265 285

FT TRANSMEM 487 507

FT TRANSMEM 515 535

FT TRANSMEM 537 557

FT TRANSMEM 564 584

FT TRANSMEM 598 618

FT TRANSMEM 671 683

FT MOD\_RES 271 271

FT MOD\_RES 683

PHOSPHORYLATION (BY SIMILARITY).

PHOSPHORYLATION (BY SIMILARITY).

FT MOD\_RES 847 847 PHOSPHORYLATION (BY SIMILARITY).

SQ SEQUENCE 940 AA; 101619 MW; 44B836307FDA36EF CRC64;

Query Match 7.5%; Score 77.5; DB 1; Length 940;

Best Local Similarity 26.1%; Pred. No. 27;

Matches 48; Conservative 22; Mismatches 55; Indels 59; Gaps 11;

QY 30 HRFARP---BALGGLTHTIAGDLADDFISLLENELKRVASINAAQNPNNVYLLP 86

DB 272 HAFVAPLWYTSAGSGLQEIYVQVWIRPDPHLSN-----YVARVIGWVDPNNSMVIIP 326

QY 87 QAI--QVGSPTH-----PSYSGHATQNGAPATVLKALIGLDRGCECFNPVPF 133

DB 327 GALLNGQSGSTGMTSKDNGALSAYMSKST---AFLT-WKDLV----- 367

QY 134 SDGGLLEINFEAGACLYEVEIKLAVN-VAFGRQMLGHYRFQIQG--LILGE---TI 186

DB 368 --DG-----LTFKGNFDMKWAENGLDGSNKIWLGLNGSILGKLLSDGNVYTI 415

QY 187 TVRT 190

DB 416 TFKT 419

RESULT 13

ID ECOLI

TO LC ECOLI STANDARD; PRT; 495 AA.

AC P02930;

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-MAY-1991 (Rel. 18, Last sequence update)

DE Outer membrane protein tolc precursor.

GN TOLC'OR MTCB OR MUKA OR REFI OR B3035.

OS Escherichia coli.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI\_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=K12;

RX MEDLINE=91016844; PubMed=2216730;

RA Niki H., Imamura R., Ogura T., Hiraga S.;

RT "Nucleotide sequence of the tolc gene of Escherichia coli.";

RL Nucleic Acids Res. 18:5547-5547 (1990).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=84015386; PubMed=6312426;

RA Hackett J., Reeves P.;

RT "Primary structure of the tolc gene that codes for an outer membrane protein of Escherichia coli K12.";

RL Nucleic Acids Res. 11:6487-6495 (1983).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474 (1997).

RN [4]

RP SEQUENCE OF 25-36.

RC STRAIN=K12 / EMG2;

RX MEDLINE=97443975; PubMed=9298646;

RA Link A.J., Robison K., Church G.M.;

RT "Comparing the predicted and observed properties of proteins encoded in the genome of Escherichia coli K-12.";

RL Electrophoresis 18:1259-1313 (1997).

RN [5]

RP SEQUENCE OF 25-29.

RC STRAIN=K12 / W3110;

RX MEDLINE=98291876; PubMed=9629924;

RA Molloy M.P., Herbert B.R., Walsh B.J., Tyler M.I., Traini M.,  
RA Sanchez J.-C., Hochstrasser D.F., Williams K.L., Gooley A.A.,  
RT "Extraction of membrane proteins by differential solubilization for  
RL separation using two-dimensional gel electrophoresis.",  
RN Electrophoresis 19:837-844(1998).  
[6]  
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 25-452.  
RX MEDLINE=20335872; PubMed=10879525;  
RA Koronakis V., Sharif A., Koronakis E., Luisi B., Hughes C.;  
RT "Crystal structure of the bacterial membrane protein TolC central to  
RT multidrug efflux and protein export.",  
RL Nature 405:914-919(2000).  
CC -1- FUNCTION: REQUIRED FOR PROPER EXPRESSION OF OUTER MEMBRANE PROTEIN  
CC COLICIN E1. MAY BE SPECIALIZED FOR SIGNAL SEQUENCE INDEPENDENT,  
CC EXTRACELLULAR SECRETION IN GRAM-NEGATIVE BACTERIA.  
CC -1- SUBUNIT: Homotrimer that assembles to form a continuous, solvent-  
CC accessible conduit: a 'channel-tunnel' over 140 Angstroms long  
CC that spans both the outer membrane and periplasmic space. The  
CC periplasmic or proximal end of the tunnel is sealed by sets of  
CC coiled helices.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.  
CC -1- SIMILARITY: BELONGS TO THE PRF FAMILY OF SECRETION PROTEINS.  
CC -1- CAUTION: REF.2 SEQUENCE WAS EXTENSIVELY CORRECTED BY REF.1.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X54049; CAA37982.1; --  
DR EMBL; X00016; CAA24914.1; ALT\_SEQ.  
DR EMBL; V01505; CAA24751.1; --  
DR EMBL; U28377; AAA69203.1; --  
DR EMBL; AE000385; AAC76071.1; --  
DR F1R; A65091; MMECTC.  
DR PDB; 1EK9; 28-JUN-00.  
DR EcoGene; EGI1009; tolC.  
DR InterPro; IPR003423; OEP.  
DR Pfam; PF02321; OEP; 2.  
KW Transport; Membrane; Outer membrane; Transmembrane; Signal; Repeat;  
KW 3D-structure; Complete proteome.  
FT SIGNAL 1 24  
FT CHAIN 25 495 OUTER MEMBRANE PROTEIN TOLC.  
FT DOMAIN 25 64 PERIPLASMIC.  
FT TRANSMEM 65 76 S1  
FT DOMAIN 77 84 EXTRACELLULAR.  
FT TRANSMEM 85 98 S2.  
FT DOMAIN 99 270 PERIPLASMIC.  
FT TRANSMEM 271 281 S4.  
FT DOMAIN 282 302 EXTRACELLULAR.  
FT TRANSMEM 303 313 S5.  
FT DOMAIN 314 495 PERIPLASMIC.  
FT REPEAT 25 232 1.  
FT REPEAT 233 448 2.  
FT CONFLICT 193 193 L -> V (IN REF. 3).  
FT HELIX 27 37  
FT HELIX 39 59  
FT HELIX 60 62  
FT STRAND 65 77  
FT TURN 82 83  
FT STRAND 85 100  
FT HELIX 102 169  
FT TURN 170 171  
FT HELIX 175 210  
FT STRAND 216 220  
FT TURN 222 224  
FT HELIX 233 243  
FT HELIX 245 265

FT HELIX 266 268  
FT STRAND 271 281  
FT TURN 284 286  
FT TURN 288 289  
FT TURN 297 298  
FT STRAND 304 318  
FT TURN 319 320  
FT HELIX 321 387  
FT TURN 388 389  
FT HELIX 393 428  
FT TURN 429 429  
FT HELIX 433 441  
FT TURN 442 442  
FT STRAND 443 448  
SQ SEQUENCE 495 AA; 54014 MW; 2AC6C964C91FA645 CRC64;  
Query Match 7.4%; Score 77; DB 1; Length 495;  
Best Local Similarity 24.4%; Pred.No.13;  
Matches 38; Conservative 21; Mismatches 57; Indels 40; Gaps 6;  
QY 11 LIQAASLAQASCYQKQVHRFARPEALGGTLHTNTAGLDADFDISLLENDELLKVAE 70  
DB 15 LSGFSLSQAEINLMQVYQARLSNP-----RKSAAADRDAFE-----K 54  
QY 71 INAAQPNNEVTLLPQAIQVGSPTHPSPSGHATQNGAFATVTKALIGLDRCGCFPNP 130  
DB 55 INEARGP-----LLP---QLGLGADYTVSGYRDANGINSNATSASLQTQS----- 98  
QY 131 VFPSDGLLEINPEGAC----LTYEGEINKLAVNVA 162  
DB 99 IFDMXKWRALTQLEKAAAGIQDVTYQTDQDTLLNTA 134  
RESULT 14  
PGK\_BACME STANDARD; PRT; 394 AA.  
ID PGK\_BACME  
AC P24269;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Phosphoglycerate kinase (SC 2.7.2.3).  
GN PGK.  
OS Bacillus megaterium.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1404;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DSM 319;  
RX MEDLINE=91057129; PubMed=2123031;  
RA Schlaepfer B.S., Branlant C., Branlant G., Zuber H.;  
RT "Nucleotide sequence of the phosphoglycerate kinase gene from  
RT Bacillus megaterium";  
RL Nucleic Acids Res. 18:6423-6423(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DSM 319.  
RX MEDLINE=93083995; PubMed=1452037;  
RA Schlaepfer B.S., Zuber H.;  
RT "Cloning and sequencing of the genes encoding  
RT glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate kinase and  
RT triosephosphate isomerase (gap operon) from mesophilic Bacillus  
RT megaterium: comparison with corresponding sequences from thermophilic  
RT Bacillus stearothermophilus";  
RL Gene 122:53-62(1992).  
CC -1- CATALYTIC ACTIVITY: ATP + 3-phospho-D-glycerate = ADP + 3-  
CC phospho-D-glyceroyl phosphate.  
CC -1- PATHWAY: Second phase of glycolysis; second step.  
CC -1- SUBUNIT: Monomer.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: Belongs to the phosphoglycerate kinase family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/send an email to license@isb-sib.ch>).

DR	ENBL; U15182; AAAC2998.1; ALT_INIT.	
DR	ENBL; Z95151; CAB08404.1; ALT_INIT.	
DR	ENBL; AL583924; CAC31168.1; ALT_INIT.	
DR	Leptoma; ML2213; -	
DR	MEROPS; M18. UPW; -	
DR	HAMAP; MF 00467; -; 1.	
DR	InterPro; IPR001948; Peptidase M18.	
DR	Pfam; PF02127; Peptidase M18; 1.	
DR	PRINTS; PR00932; AMINOPEPTASE.	
KW	Hypothetical protein; Hydrolase; Aminopeptidase; Metalloprotease;	
KW	Zinc; Complete proteome.	
FT	METAL	79
		ZINC (POTENTIAL) .

FI	METAL	156	ZINC (POTENTIAL).
FT	METAL	399	ZINC (POTENTIAL).
SQ	SEQUENCE	426 AA; 45593 MW; 582ADDD29C3A0A53 CRC64;	

[illegible]

```

Db      323  --LIEVNAGPVLKVHENLRY-ATDGTAAAFEVAC-----QJAGVRL 361
QY      171  ---HYRFDGIQGLLGETITVRT 190
Db      362  QRYEHEADRPCGSTIGFLASART 384

```

Search completed: September 19, 2004, 03:55:58  
Job time : 9.02288 secs

Search completed: September 19, 2004, 03:55:58  
Job time : 9.02288 secs

***This Page Blank (uspto)***



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 19, 2004, 03:59:53 ; Search time 5444.35 Seconds  
(without alignments)  
3707.849 Million cell updates/sec

Title: US-10-691-383-2  
Perfect score: 3528  
Sequence: 1 MLCXAADTRGSPMPDTGVL.....SIDGDMCSGLVYTGVDACQA 676

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2 1/USPTO.spool/US10691383/runat 17092004 102656 1753/app query.fasta\_1.1230  
-DB=EST -QPM=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=dlosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10691383 @CGN 1 1 4591 @runat 17092004 102656 1753 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vri:\*  
28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	217	6.2	393	10	AW400475	AW400475 Lamdisest
2	199.5	5.7	503	9	AU187248	AU187248 Lamdisest
3	199.5	5.7	538	9	AU191942	AU191942 Lamdisest
4	146	4.1	503	9	AU194816	AU194816 Lamdisest
5	138.5	3.9	441	9	AU187606	AU187606 Lamdisest
6	117	3.3	1201	9	AL539002	AL539002 Lamdisest
7	115	3.3	2089	29	AY411102	AY411102 Homo sapi
8	115	3.3	2673	11	BC040379	BC040379 Homo sapi
9	114.5	3.2	889	12	BI836552	BI836552 603089514
10	113	3.2	369	14	CD063088	CD063088 MA1-0029U
11	113	3.2	936	11	CNS080UB	BX007359 Single re
12	110.5	3.1	701	14	CB552155	CB552155 MMSPP0080
13	110.5	3.1	2823	11	AK039670	AK039670 Mus muscu
14	110	3.1	1052	29	CNS03HC2	AL244091 Tetraacu
15	109	3.1	712	14	CB443557	CB443557 694466 MA
16	108	3.1	392	9	AU189808	AU189808 Lamdisest
17	108	3.1	1168	12	BM456445	BM456445 AGENCOURT
18	106.5	3.0	668	28	BH86578	BH86578 LB00877A
19	105.5	3.0	3208	11	AK039695	AK039695 Mus muscu
20	105	3.0	3250	11	BC047878	BC047878 Homo sapi
21	104	2.9	3371	29	AY398790	AY398790 Homo sapi
22	103.5	2.9	995	14	CF224754	CF224754 AGENCOURT
23	103.5	2.9	1680	12	BM925460	BM925460 AGENCOURT
24	103.5	2.9	3254	11	CNSLT170	EX161390 human ful
25	103	2.9	3273	29	AY398791	AY398791 Pan trogl
26	102	2.9	3116	11	AK045649	AK045649 Mus muscu
27	102	2.9	3117	11	AK045649	AK045649 Mus muscu
28	101.5	2.9	852	28	BZ709996	BZ709996 OGBAR23TC
29	101.5	2.9	1050	12	BM805002	BM805002 AGENCOURT
30	101	2.9	1050	12	BM805002	BM805002 AGENCOURT
31	100.5	2.8	844	14	CD519980	CD519980 AGENCOURT
32	100.5	2.8	1323	28	BZ572188	BZ572188 msn2_2542
33	100	2.8	400	10	AW400432	AW400432 Lamdisest
34	100	2.8	787	28	BZ710005	BZ710005 OGBAR23TM
35	99.5	2.8	780	29	CC700975	CC700975 OGUFW58TH
36	99.5	2.8	1398	29	U90627	U90627 U90627 Azos
37	99	2.8	655	10	BE822718	BE822718 GM700018B
38	99	2.8	913	12	BG741433	BG741433 602632215
39	99	2.8	1054	12	BI414575	BI414575 602986154
40	98.5	2.8	584	12	BG897929	BG897929 HOA2-G-1
41	98.5	2.8	930	13	BQ939745	BQ939745 AGENCOURT
42	98.5	2.8	1027	29	CG687286	CG687286 ZMMBB016
43	98.5	2.8	1125	10	BE307732	BE307732 601095594
44	98.5	2.8	1504	11	AY104439	AY104439 Zee may
45	98	2.8	754	13	BU100257	BU100257 WHE3351_B

ALIGNMENTS

RESULT 1  
AW400475  
LOCUS  
DEFINITION  
Lamdisest169est L.digitata sporophyte Lambda ZapII Laminaria  
digitata cDNA similar to vanadium bromoperoxidase, mRNA sequence.  
ACCESSION  
AW400475  
VERSION  
AW400475.1  
KEYWORDS  
EST.  
SOURCE  
Laminaria digitata  
ORGANISM  
Laminaria digitata  
Eukaryota; stramenopiles; Phaeophyceae; Laminariales;  
Laminariaceae; Laminaria.  
REFERENCE  
1 (bases 1 to 393)



```

ACCESSION   AU191942
VERSION     AU191942.1
KEYWORDS    GI:31930088
SOURCE      EST.
ORGANISM    Porphyra yezoensis
            Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
            Porphyra.
REFERENCE   1 (bases 1 to 538)
AUTHORS     Asamizu,B., Nakajima,M., Kitade,Y., Saga,N., Nakamura,Y. and
            Tabata,S.
TITLE       COMPARISON OF RNA EXPRESSION PROFILES BETWEEN THE TWO GENERATIONS
            OF PORPHYRA YEZOENSIS (RHODOPHYTA), BASED ON EXPRESSED SEQUENCE TAG
            FREQUENCY ANALYSIS
JOURNAL     J. Phycol. 39 (5), 923-930 (2003)
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES    Location/Qualifiers
             source
               1..538
               /organism="Porphyra yezoensis"
               /mol_type="mRNA"
               /strain="TU-1"
               /db_xref="taxon:2788"
               /clone="PFL023f08_r"
               /dev_stage="sporophytes"
               /clone_lib="Porphyra yezoensis TU-1 sporophytes"
ORIGIN
Alignment Scores:
Pred. No.:      2.84e-10      Length:      538
Score:          199.50      Matches:    57
Percent Similarity: 51.54%      Conservative: 10
Best Local Similarity: 43.85%      Mismatches: 54
Query Match:    5.65%      Indels:    9
DB:              9      Gaps:      4

US-10-691-383-2 (1-676) x AU191942 (1-538)
QY  429 AspArgGlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeuLeu---Gly 447
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  91 GATCATCAGCAGGGTTTGCCAGTGTGGGGGCCCCCACCCTGTGACGCTGCACCGAG 150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  448 AlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAla 467
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  151 GTGCCACACGCGCACTCAAGCGGTGCGCTTCCAAAGTTCAACGTCCACCGCGGCTC 210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  468 ArgProGluAlaLeuGlyThrLeuHisAsn---ThrIleAlaGlyAspLeuAspAla 486
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  211 CGGCCCGAGCGGTGGCGGCGCTCTCTCGACCGGCTGCGCGCCGCGCAGGACCGCTGGCG 270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  487 AspPheAspIleSerLeuLeuGluAsnAspGluLeuLysArgValAlaGluIleAsn 506
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  271 CCAGTGGCCCGCTGGCGAGAGGAGCGCGGAGACGCTCGCGCGGGGTGTGATGCCAAC 330
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  507 AlaAla-----GlnAsnProAsnGluValThrTyr-----LeuLeu 519
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  331 GCTGCCCGCCGCTGGCGGGGCGAGTGGCGCGGTGGCGCGGCTGTACGACAGTGTGTGCTG 390
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  520 ProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThr 539
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  391 CCAATGGCCCTCCAGAGGGGCTCCCGGATGCACCCCTCTCTACGGCGCGGTACGCCACC 450
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  540 GlnAsnGlyAlaPheAlaThrValLeuLys 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  451 GTTGGCGGCGCGTGGCGGTGCGTCAAG 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
AU194816
LOCUS      AU194816 Porphyra yezoensis TU-1 sporophytes Porphyra yezoensis
DEFINITION CDNA clone PFL065g09_r 5', mRNA sequence.

AU194816 503 bp mRNA linear EST 14-OCT-2003

```

```

ACCESSION   AU194816
VERSION     AU194816.1
KEYWORDS    GI:31935825
SOURCE      EST.
ORGANISM    Porphyra yezoensis
            Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
            Porphyra.
REFERENCE   1 (bases 1 to 503)
AUTHORS     Asamizu,B., Nakajima,M., Kitade,Y., Saga,N., Nakamura,Y. and
            Tabata,S.
TITLE       COMPARISON OF RNA EXPRESSION PROFILES BETWEEN THE TWO GENERATIONS
            OF PORPHYRA YEZOENSIS (RHODOPHYTA), BASED ON EXPRESSED SEQUENCE TAG
            FREQUENCY ANALYSIS
JOURNAL     J. Phycol. 39 (5), 923-930 (2003)
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES    Location/Qualifiers
             source
               1..503
               /organism="Porphyra yezoensis"
               /mol_type="mRNA"
               /strain="TU-1"
               /db_xref="taxon:2788"
               /clone="PFL065g09_r"
               /dev_stage="sporophytes"
               /clone_lib="Porphyra yezoensis TU-1 sporophytes"
ORIGIN
Alignment Scores:
Pred. No.:      0.000114      Length:      503
Score:          146.00      Matches:    42
Percent Similarity: 49.52%      Conservative: 10
Best Local Similarity: 40.00%      Mismatches: 43
Query Match:    4.14%      Indels:    10
DB:              9      Gaps:      3

US-10-691-383-2 (1-676) x AU194816 (1-503)
QY  177 PheHisIysSerLeuProHisAspGluLeuGlyGlnValThrAlaAspAsp-----Phe 194
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  48  TTCAACAGGGTCTGCAGCACACCTCTCGACGGGCTCTTGGAGGACCCCGCGCCTTT 107
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  195 AlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAspValProAla 214
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  108 CGCGCCTTTGTCGGGCCATTGACAGCGGGGACCCGCGGAGCTTTGTTCGACGTGCCCAT 167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  215 GlyAsp-----ProAlaGlyArgLeuVal-----AsnProThrAla 226
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  168 GCGGACCGCTTACCGCGCCGCCACCCCGGGGTGCGGTGCGGAGTCGCGCTCGCGC 227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  227 AlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThrThrIleProProValPro 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  228 GGCGTGGCGTTTACCTTCAGGGCCCCGATGCGCAGCGCTGACCATGCCCGCCCGCCG 287
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  247 ThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAla 266
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  288 CGGCTGGGACGCGGNACTGTGTGGCGGAGATGATTGAGGTCTACGCCATGGCGCTCGCG 347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  267 ArgAspValProPhe 271
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  348 CGGGATGTGCGGTTT 362
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
AU187606
LOCUS      AU187606 Porphyra yezoensis TU-1 sporophytes Porphyra yezoensis
DEFINITION CDNA clone PFL014h1_r 5', mRNA sequence.

AU187606 441 bp mRNA linear EST 14-OCT-2003

```

**ORGANISM** Porphyra yezoensis  
Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;  
**REFERENCE** Porphyra.  
**AUTHORS** 1 (bases 1 to 441)  
Asamizu,E., Nakajima,M., Kitade,Y., Saga,N., Nakamura,Y. and  
Tabata,S.:  
**TITLE** COMPARISON OF RNA EXPRESSION PROFILES BETWEEN THE TWO GENERATIONS  
OF PORPHYRA YEZOENSIS (RHODOPHYTA), BASED ON EXPRESSED SEQUENCE TAG  
FREQUENCY ANALYSIS  
**JOURNAL** J. Phycol. 39 (5), 923-930 (2003)  
**COMMENT** Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/  
**FEATURES** Location/Qualifiers  
**source**  
1..441  
    /organism="Porphyra yezoensis"  
    /mol\_type="mRNA"  
    /strain="TU-1"  
    /db\_xref="taxon:2788"  
    /clone="PF014hl1\_r"  
    /dev\_stage="sporophytes"  
    /clone\_lib="Porphyra yezoensis TU-1 sporophytes"

**ORIGIN**

Alignment Scores:			
Pred. No.:	0.000559	Length:	441
Score:	138.50	Matches:	43
Percent Similarity:	43.97%	Conservative:	19
Best Local Similarity:	30.50%	Mismatches:	72
Query Match:	3.93%	Indels:	7
DB:	9	Gaps:	3

US-10-691-383-2 (1-676) x AU187606 (1-441)

```

Qy      336 ThrIleAaspAlaileThrValGluProLyseGlInlThrPheAlaProAaspLeuAasnTyr 355
          ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      17  ACGTAGCGGCCAACCGGGTAGACCTCGGGTGCCCGTTGCAGCGCGGGGTGTCGACTAC 76

Qy      356 MetValaspPheaspGluThrLpueAasnIlleGlnAasnGlyGlyProProAlaglyProGlu 375
          ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db      77  CTCACCAGCTGGGAGTGTGTTGTTCGAGCTCCAAAACGGCGCAGACACCCCGCGCACCGGAG 136

Qy      376 Gluleu-----AaspGluGluleuArgPheilleArgasnalAargAaspLeu 390
          ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db      137 ACGTAGCTGCCCGCGGGATGGACCAAAACCGCTTCATTGCCACCCCGCGGACCTT 196

Qy      391 AlaArgValSerPheValaspAasnilleAasnThrGluAlaTyrArgGlySerLeuilleLeu 410
          ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      197 GCCACGTACCTCCACTACGATGCTCTGTACCAGGCATACCTTGAACGGGCTCTGTGGCTG 256

Qy      411 LeuGluLeuGlyAlaPheSerArgProGlyIyleAasnGlyProPheilleAaspSer---Aasp 429
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      257 CTAGCGAGGGTGTGCCCTTTTGACCGCGGCTCTCCGTGGAGCCGCCCGCCGACGAGGTGGAT 316

Qy      430 ArgGlnAlaGlyPheValAasnPheGlyThrSerHisTyrPheargLeulle---GlyAla 448
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      317 CATCAGCAGGGGTTTGCACGTTTGGGGCCCCCACCTGTGACGTCTGCACCGGAGGTG 376

Qy      449 AlaGluLeuAlaGlnArGalaserCyserTyrGlnlystirpGlnValHisArgPheAlaArg 468
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      377 GCACACCGCGCACTCAGGGGGTGGCGTTCAAAAAGTTCAACGTCCACCGCCGCTCCGG 436

Qy      469 Pro 469
          |||||
Db      437 CCC 439
    
```

**RESULT 6**  
AL539002  
**LOCUS** AL539002 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone  
**DEFINITION** CS0DF030Y123 5'-PRIME. mRNA sequence.



```
Qy 348 ThrPheAlaProAspLeuAsnTyrMetValaspPhe-----AspGluTyrLeuAsn 364
Db 716 AAACCTTCTAAGTAT---GGCTATGCTGTGAAGTAGATGCCAAATATCTACTACT 772
Qy 365 IleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluGluLeuArgPheIle 384
Db 773 ACAGTTATGCTGGCAGCAAGATGGTCCAGGATCGAGCTACAGAGCTAGCT---GTA 829
Qy 385 ArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsnThrGluAlaTyr 404
Db 830 AAATCTTTCAGAAACGCTATTGAAGTTAGAGTTGTCCGGCGGAAATTCAGAGCATACATTT 889
Qy 405 ArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSer-----ArgProGlyIle 421
Db 890 GGAGAGAGACCTCGGTGGTGTGCATATGGCATGTCTCTCATGATTTCTCCAGCTCTG 949
Qy 422 AsnGlyPro-----PheIleAspSerAspArgGlnAlaGlyPheVal 435
Db 950 AAGCCACCAGACGACCTGGTGGTGAAGATGCTTCGGGCGACCCCAATTTT--- 1006
Qy 436 AsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAla 455
Db 1007 AATGCTTCTGCCAAACACTGGACCAATTTTGTCT-----ATTACAGAAATGCA 1054
Qy 456 SerCysTyrGlnLysTyrGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThr 475
Db 1055 AAT-----GATGCAATTTGGTATCCTT 1075
Qy 476 LeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsn 495
Db 1076 AACAAATTCCTCCATTCACACAGATGCTCAATTGAATCAAAATATGATGATGCCAAAT 1135
Qy 496 -----AspGluLeuLeuLys-ArgValAlaGluIle 505
Db 1136 CGCACATGGCGTGTGCGAGTGTCTTTTACAAGATCACTGTCTAGTGAAGGTATGGAA-- 1193
Qy 505 eAsnAlaAlaGlnAsnProAsnGluValThrTyrLeuLeuPro--GlnAlaIleGln 524
Db 1194 -----CCAGAAACCAAGTAATACATCGACGTCCGACGAGGCTTTGAAA 1237
Qy 525 ValGlySerProThrHisProSerTyrProSer-----Gly 536
Db 1238 ATTCTTCAAAAACACACAGCCCACTTATCCATCTGTCAAAAGTTCCACAATGCCATACAGGC 1297
Qy 537 HisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArg 556
Db 1298 TCTTCAACCCAGAGGATCTGGAAGAAGAAAGATATAAGGATCTGTAGTTATGAGAAT 1357
Qy 557 GlyGlyGlyCysPheProAsnProValPheProSerAspGlyLeuGluLeuLeuAsn 576
Db 1358 TCTTCA-----AATCCGTTGCGCGCTGAACGAC----- 1387
Qy 577 PheGluGlyAlaCysLeuThrTyrGluGlyLeuIleAsnLysLeuAlaValAsnValAla 596
Db 1388 -----ACAGCTCAGTTTAAACCAATGACAGTTGATGATGTC 1423
Qy 597 PheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIle-----GlnGlyLeuLeu 615
Db 1424 TATGAAAGG---ATGACAGCGCTCCGCTGGAAATGCAAGATGATTCTTAGAGAGTGAAGTA 1480
Qy 616 LeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluGlu 635
Db 1481 ATTGCAAGACAGTGTGGGTGAAG-----AAAACGTCAATATGAA 1522
Qy 636 AlaThrPheGluPheArgLeuPheThrGlyGluValIleLysLeuPheGlnAspGlyThr 655
Db 1523 GCT-----GCTGGGAAGCTGTGAAACCCCTCAAAAGACCCAG 1561
Qy 656 PheSerIleAspGlyAspMetCysSerGlyLeuVal 667
Db 1562 CCAACTGTCTTAACAACCTTGAAGAAAGGAGCTGTT 1597
```

## RESULT 8

BC040379 2673 bp mRNA linear HTC 19-NOV-2003  
Homo sapiens NF-kappa B-repressing factor, mRNA (cdna clone  
IMAGE:477539), containing frame-shift errors.

BC040379

HTC.

BC040379.1 GI:25955669

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2673)

REFERENCE

AUTHORS

Klausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, P.S., Wagner, L., Shennen, C.M., Schuler, G.D.,  
Altschul, S.F., Zebberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toohiyuki, S.,  
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
Adams, R.D., Mullany, S.J., Bosak, S.A., McSwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Sanchez, A., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,  
Bouffard, G.C., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzyzinski, M.I., Skalska, U., Smailus, D.E.,  
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length

human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

12477932

REFERENCE

2 (bases 1 to 2673)

AUTHORS

TITLE

JOURNAL

COMMENT

REMARK

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 29 Row: 1 Column: 16  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 8923943  
This clone has the following problem: frame shifted.

Location/Qualifiers

1. .2673

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:477539"

/tissue\_type="Skin, normal"

/clone\_lib="NCI CGAP\_Skn3"

/lab\_host="DH10B"

/note="Vector: pCMV-SPORT6.1"

## ORIGIN

## Alignment Scores:

Pred. No.: 4.01 Length: 2673  
Score: 115.00 Matches: 134  
Percent Similarity: 34.43% Conservative: 96  
Best Local Similarity: 20.06% Mismatches: 260  
Query Match: 3.26% Indels: 179  
DB: 11 Gaps: 34

US-10-691-383-2 (1-676) x BC040379 (1-2673)

Qy 66 LeuProThrAspGlyIleSerAlaSerIleLeuGlyIleMetAlaArgValArg 85  
Dy |||||  
Qy 72 CTGCGCTCGACCGCTCTCGCGCTCTCGCGCGCTCGACCAACCGCTCTCTCGGGCT 131  
Dy |||||  
Qy 86 IleAlaThrAla-----LeuAlaValLeuAla 95  
Dy |||||  
Dy 132 GCAGGTACAGCCACCGCTTGATGGAAAAAATTCCTCAAATGGCTGAAGGTATTGATATTG 191  
Qy 96 AlaProCysLeuAlaPheAspGluValThrAlaSerGlyValPheProGluGluHisLys 115  
Dy |||||  
Dy 192 GGGAGATGCCTTCA-TATGATCTGGTGTCTGTCAAA-----CCTTCCAAA----- 235  
Qy 116 HisThrGlyGluGlyArgHisLeuGlnThrCysThrAsnSerAspAlaLeuAspPro 135  
Dy |||||  
Dy 236 -----GGTCAAAACGCCCTCTCAACATGT-----GATGGTCAAAATCCT 277  
Qy 136 Thr-----AlaProAsnArgArgAspAsnValAlaPhe 146  
Dy |||||  
Dy 278 CTTAAAAAGCAAGCGGTTCCAAATTCATGCGAGACCTCGTTTGAGCCTGTACATTTT 337  
Qy 147 AlaSerArgArgAspAlaAlaArgArgGluArgAspGlyThrGlyThrValCysGlnIle 166  
Dy |||||  
Dy 338 GTAGCTAGTAGTTCAAAGATGAAGACAGAGAGATCCTTATGGCCCTCAACAAAGAG 397  
Qy 167 ThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSerLeuProHisAspGluLeu 186  
Dy |||||  
Dy 398 GTAAATGAACAACACATTTTGCC-----AGCATGCCAAGACACATCTAC 442  
Qy 187 GlyGlnValThrAlaAspAspPheAlaIleLeuGluAsp-----CysIleLeu 202  
Dy |||||  
Dy 443 CAAGATTATATCTCAAGACTCTTTTCAAGTATCAAGATGGGAATTCATGATTTGTGATTCA 502  
Qy 203 AsnGlyAspPheSerIleCysGluAspValProAlaGlyAspProAlaGlyArgLeuVal 222  
Dy |||||  
Dy 503 TCAGGA---TTCATCTCAAAAGACACAGCCTCTA----- 535  
Qy 223 AsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThr----- 240  
Dy |||||  
Dy 536 -----ACAGCCACATGATTTTGCAGTGGGAACCTGCCCAAGACACACATCACAG 589  
Qy 241 -----ThrIleProValProThrLeuSerSerProGluLeuAla 254  
Dy |||||  
Dy 590 CAGCAAACTCTCAGTCAACTCTCGAGCCTTCAACATCACAGACATTTCCCGAG----- 643  
Qy 255 AlaGlnLeuAlaGluLeuThrTrpMetAlaLeuAlaArgAspValProPheMetGlnThr 274  
Dy |||||  
Dy 644 TCTGTGTAGCCGAG-----AAGCAGTAT 667  
Qy 275 GlyThrAspGluIleThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPhePro 294  
Dy |||||  
Dy 668 TTATTGAAAAAATAACGGCGACATCTGGNAGAACCTTTCTAAT-----CCA 715  
Qy 295 AsnLeuAspAlaValSerIleGlySerAspGlyThrValAspProPheSerGlnLeuPhe 314  
Dy |||||  
Dy 716 GAAATGACTTCT-----GGATCTGATAAAAT-----AATTATACATATATGTTA 760  
Qy 315 ArgAlaThrPheValGlyValGluThrGlyPro---PheValSerGlnLeuLeuValAsn 333  
Dy |||||  
Dy 761 ACTCGTGTATTACGGCGTGTAAAGACAAATCCTCGATATATATATGCTCTTAAAGGAA 820

Qy 334 SerPheThrIleAspAlaIleThrValGluProLysGlnGluThrPheAlaProAspLeu 353  
Dy |||||  
Dy 821 ATTCTCTCTCGCGACATC-----CCCAAAATAAAACTTCTAACTGAT--- 865  
Qy 354 AsnTyrMetValAspPhe-----AspGluTrpLeuAsnIleGlnAsnGlyGlyPro 370  
Dy |||||  
Dy 866 GGCTATGCTTGTGAAGTTAGATGCCAAATATCTACTTAACACAGGTATGCTGGCAGC 925  
Qy 371 ProAlaGlyProGluGluLeuAspGluGluArgPheIleArgAsnAlaArgAspLeu 390  
Dy |||||  
Dy 926 AAGAAATGGTCCAGGATCGAGCTACAGAGCTAGCT---GTAAAACTCTTCAGAGAAACGT 982  
Qy 391 AlaArgValSerPheValAspAsnIleAsnThrGluAlaTyrArgGlySerLeuLeuLeu 410  
Dy |||||  
Dy 983 ATTGAAGTTAGAGTTGTCCGCGGAAATTCAGCATACATTTTGGAGAGGACCTCGTGTGT 1042  
Qy 411 LeuGluLeuGlyAlaPheSer-----ArgProGlyIleAsnGlyPro----- 424  
Dy |||||  
Dy 1043 TGTCAAGATTGGCATGCTCTCTATGAATTTCTCCAGTCTGAAGCCACCAAGAGACCTG 1102  
Qy 425 -----PheIleAspSerAspArgGlnAlaGlyPheValAsnPheGlyThrSerHis 441  
Dy |||||  
Dy 1103 GTGGTGTCTGGTAAAGATGCTTCTGGCAGCCCAATTTT---AATGCTTCTGCCCAACAC 1159  
Qy 442 TyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrp 461  
Dy |||||  
Dy 1160 TGGACCAATTTTGTG-----ATTACAGAAATGCAAT----- 1192  
Qy 462 GlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAla 481  
Dy |||||  
Dy 1193 -----GATGCAATTTGGTATCTCTTAACAATTTGCTCTCATTC 1228  
Qy 482 GlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsn----- 495  
Dy |||||  
Dy 1229 AACAAAGATGTCATTTGAATACAAATATGAGATGATGCCAAATCCACATGCGGTGTGCGA 1288  
Qy 496 -----AspGluLeuLeuLys-ArgValAlaGluIleAsnAlaAlaGlnAsnPr 511  
Dy |||||  
Dy 1289 GTGTTTTTACAAGATCACTGCTTAGCTGAAGGTTATGGAA-----CC 1330  
Qy 511 oAsnAsnGluValThrTyrLeuLeuPro--GlnAlaIleGlnValGlySerProThrHis 530  
Dy |||||  
Dy 1331 AAGAAACAAGTAAACATGAGCTGCCGACGAGCTTTGAAATTTCTTCAAAACACAG 1390  
Qy 531 ProSerTyrProSer-----GlyHisAlaThrGlnAsnGly 542  
Dy |||||  
Dy 1391 CCACATTATCACTGTCAAAAGTTCAATGCCATACAGGCTCTTCAACCAGAGGATCT 1450  
Qy 543 AlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyCysPhePro 562  
Dy |||||  
Dy 1451 GGAAGAGAAAGAAAGATATAAGGATCTTGTAGTTTATGAGAAATTTCTTCA----- 1498  
Qy 563 AsnProValPheProSerAspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeu 582  
Dy |||||  
Dy 1499 AATCCCGTGTGCACGCTGAACGAC----- 1522  
Qy 583 ThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeu 602  
Dy |||||  
Dy 1523 -----ACAGCTCAGTTTAAACCGAATGACATGATGATGCTATGAAGG---ATGACA 1573  
Qy 603 GlyIleHisTyrArgPheAspGlyIle---GlnGlyLeuLeuLeuGlyGlyThrIleThr 621  
Dy |||||  
Dy 1574 GGCTCCGCTGGAAATGCAAAAGTGTCTAGAGAGTGAAGTAAATTCGAGAGACAGTGGG 1633  
Qy 622 ValArgThrLeuHisGlnGluLeuMetThrPheAlaGluAlaThrPheGluPheArg 641  
Dy |||||  
Dy 1634 GTGAAG-----AAAATCTGCTCAATATATGAAGCT----- 1660  
Qy 642 LeuPheThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIleAspGlyAsp 661  
Dy |||||  
Dy 1661 -----CCTGGGAGAGCTGTGAACCAACCTCAAAAGACCCCACTGCTATTAACAC 1714  
Qy 662 MetCysSerGlyLeuVal 667









Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 20530913  
 MEDLINE  
 PUBMED  
 11076861  
 4

**TITLE**  
 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

**JOURNAL**  
 Functional annotation of a full-length mouse cDNA collection

**REFERENCE**  
 Nature 409, 685-690 (2001)

**AUTHORS**  
 5

**TITLE**  
 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

**JOURNAL**  
 Analysis of the mouse transcriptome based on functional annotation

**REFERENCE**  
 of 60,770 full-length cDNAs

**AUTHORS**  
 Nature 420, 563-573 (2002)

**TITLE**  
 6 (bases 1 to 2823)

**JOURNAL**  
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,

**REFERENCE**  
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,

**AUTHORS**  
 Hayashida, K., Hayatsu, N., Hitamoto, K., Hiraoka, T., Hirozane, T.,

**TITLE**  
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,

**JOURNAL**  
 Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,

**REFERENCE**  
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,

**AUTHORS**  
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,

**TITLE**  
 Okazaki, Y., Saito, K., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,

**JOURNAL**  
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,

**REFERENCE**  
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,

**AUTHORS**  
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,

**TITLE**  
 Muramatsu, M. and Hayashizaki, Y.

**JOURNAL**  
 Submitted

**REFERENCE**  
 Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of

**AUTHORS**  
 Physical and Chemical Research (RIKEN), Laboratory for Genome

**TITLE**  
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),

**JOURNAL**  
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,

**REFERENCE**  
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,

**AUTHORS**  
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,

**TITLE**  
 Fax: 81-45-503-9216)

**JOURNAL**  
 A cDNA library was prepared and sequenced in Mouse Genome

**REFERENCE**  
 Encyclopedia Project of Genome Exploration Research Group in Riken

**AUTHORS**  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.

**TITLE**  
 Division of Experimental Animal Research in Riken contributed to

**JOURNAL**  
 prepare mouse tissues.

**REFERENCE**  
 Please visit our web site for further details.

**AUTHORS**  
 URL: http://genome.gsc.riken.go.jp/

**TITLE**  
 URL: http://fantom.gsc.riken.go.jp/.

**JOURNAL**  
 Location/Qualifiers

**REFERENCE**  
 1. .2823

**AUTHORS**  
 /organism="Mus musculus"

**TITLE**  
 /mol\_type="mRNA"

**JOURNAL**  
 /strain="C57BL/6J"

**REFERENCE**  
 /db\_xref="FANTOM DB:A330082D18"

**AUTHORS**  
 /db\_xref="MGI:2404165"

**TITLE**  
 /db\_xref="taxon:10090"

**JOURNAL**  
 /clone="A330082D18"

**REFERENCE**  
 /sex="male"

**AUTHORS**  
 /tissue types="spinal cord"

**TITLE**  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"

**JOURNAL**  
 /dev\_stage="adult"

**REFERENCE**  
 1. .2823

**AUTHORS**  
 /notes="hypothetical Serine-rich region/IPT/TIG

**TITLE**  
 domain/Ankyrin repeat region circular profile/Yeast

**JOURNAL**  
 DNA-binding domain containing protein (PROSITE|PS50324,

**REFERENCE**  
 InterPro|IPR002909, PROSITE|PS50297, InterPro|IPR003163,

**AUTHORS**  
 evidence: InterPro")

**TITLE**  
 Alignment Scores:

**JOURNAL**  
 Pred. No.: 13.2 Length: 2823

**REFERENCE**  
 Score: 110.50 Matches: 132

**AUTHORS**  
 Percent Similarity: 34.74% Conservative: 74

**TITLE**  
 Best Local Similarity: 22.26% Mismatches: 202

**JOURNAL**  
 Mismatches: 202

**REFERENCE**  
 Mismatches: 202

**AUTHORS**  
 Mismatches: 202

**TITLE**  
 Mismatches: 202

**JOURNAL**  
 Mismatches: 202

**REFERENCE**  
 Mismatches: 202

**AUTHORS**  
 Mismatches: 202

**TITLE**  
 Mismatches: 202

**JOURNAL**  
 Mismatches: 202

**REFERENCE**  
 Mismatches: 202

**AUTHORS**  
 Mismatches: 202

Query Match: 3.13% Indels: 188  
 DB: 11 Gaps: 30  
 US-10-691-383-2 (1-676) x AK039670 (1-2823)  
 QY 162 ThrValCysGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSerLeu 181  
 DB 1167 ACCTCCTGTGAGCTTGCCGCTGTTGTAGTT-CTGAGCCTTCTGCCCCACCGGCTCTC 1255  
 QY 182 Pro-----HisAsp 184  
 DB 1226 CAGCCCTGCTTTGACCCCTGATCGTTTCTCAACGCCCCCAAGGGCCGACATACG 1285  
 QY 185 GluLeuGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCysIleLeuAsnGly 204  
 DB 1286 GAGGG-GGCCAAGAGTAAACCCAGACTTCCCGAGGACAG-----GGC 1329  
 QY 205 AspPheSerIleCysGluAspValProAlaGlyAspProAlaGlyArgLeuValAsnPro 224  
 DB 1330 ACTCACACTCCCTGT-----CCTGCCCTGGAACTCGAGCTGCCCTGGAGCCCGCAG 1380  
 QY 225 ThrAlaAlaPheAlaIle----- 230  
 DB 1381 GCGGCTGCTCGAGTCTCGCTCCACAGTTCGGAGCAACCGGAGAGGAGAAACAATTC 1440  
 QY 231 -----AspIleSerGlyProAlaPheSerAlaThrThr 241  
 DB 1441 TTTATCCAGATGATAGTGGGAGGAACCTCAAGGGTCCGGGA-----ACG 1488  
 QY 242 IleProProValProThrLeu-----SerSerProGluLeuAlaAlaGlnLeuAlaGlu 259  
 DB 1489 GTCCCGCGGTACTCTTCATCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 1548  
 QY 260 LeuTyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGluIle 279  
 DB 1548 ----- 1548  
 QY 280 ThrThrAlaAlaAlaAlaAlaAlaGlyMetGlyGlyPheProAsnLeuAlaVal 299  
 DB 1549 TCAGGCAGGGCAACGAG 1587  
 QY 300 SerIleGlySerAspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheVal 319  
 DB 1588 TCTGCTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1623  
 QY 320 GlyValGluThrGlyProPheValSerGlnLeuValAsnSerPheThrIleAspAla 339  
 DB 1624 ---CTTTCGTCATTCACAGACCTCATGGGAGAACTCATCAGTGAT-----GAAGCT 1671  
 QY 340 IleThrValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPhe 359  
 DB 1672 CCAGGTGTCTCTCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 1731  
 QY 360 Asp---GluTrpLeuAsnIleGlnAsnGlyGlyProPro-----AlaGlyPro--- 374  
 DB 1732 TCCCCAGAGTGG---TCTACCCAGAGGGCGGGGTCAAGGTGTCTCATCACAGGCCCTTGG 1788  
 QY 375 GluGluLeuAspGluGluLeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSer 394  
 DB 1789 ACAGAGCCCGCAGAGCATTTACTCTCTGTGTC----- 1818  
 QY 395 PheValAspAsnIleAsnThrGluAlaTyrArgGlySerLeuIleLeuLeuGluLeuGly 414  
 DB 1819 ---TTCGATCACATC-----GCAGTCCAGCCCTCCCTGGTC----- 1851  
 QY 415 AlaPheSerArgProGlyIleAsnGlyProPheIleAspSerAspArgGlnAlaGlyPhe 434  
 DB 1852 -----CAGCCTGTGTCTTACGCTGTCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1899  
 QY 435 ValAsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArg 454  
 DB 1900 GTGTCCTTG-----CAGGTGGCGGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1938

```
Qy 455 AlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGly 474
Db 1939 GCCTCTGCTCTTTGAGTATCGAGCGCGGTTCTCTG-----TCTCTGCCTAGT 1989

Qy 475 ThrLeuHisAsnThrIleAlaGlyAspLeuAspPheAspIleSerLeuLeuGlu 494
Db 1990 ACACAGCTCGACTGTTGTCTCACTGGAT---GACAGCCAGTTCCGATGTCATCTCGGAG 2046

Qy 495 Asn---AspGluLeuLeuLysArgValAlaGluIleAsnAlaLa----- 508
Db 2047 CGCTGGAGCAGATGGAGAGCGGATGCGAGATTCCGCGAGCTGGCGAGCTCTCGGT 2106

Qy 509 -----GlnAsnProAsnAsnGlu----- 514
Db 2107 CAGGCCCGCAGGCTCTCAATTCAGGATGAGGCGCGGCTCTGCTTCGAGCAGCG 2166

Qy 515 -----ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHis 530
Db 2167 GTGTGTGCTCTGTAGAGCATGATCCCGGTCCTCACTTG-GAGGGGTCTCAACGGCT 2225

Qy 531 ProSerTyrProSer-----GlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeu 548
Db 2226 GATCCATGGAAGTCCCTTCGCGGCGCATGAGCTCTTACATCTGCTGCTGCACAGGGCTA 2285

Qy 549 LysAlaLeu-----IleGlyLeu 554
Db 2286 CGCTCGGCTCATCGAGACTCTGAGCCAGTGGAGGATGTGGAAACAGGAGCTTGACTT 2345

Qy 555 AspArgGlyGly-----GluCysPheProAsnProValPheProSerAspAspGly 571
Db 2346 AGAGCAAGAGGTTGACCGCTCAATGTGGACCATTTCTCTTGACCCCTCTGATGTGGC 2405

Qy 572 Leu-GluLeuLeuAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnLysLe 591
Db 2406 CTGTGCTCTGGGACACTTGAAGCTGTGTGCTCTTTCTGTGGAAACGACAGCACT 2465

Qy 591 uAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIle-----HisTy 606
Db 2466 GAGCATTCGGGACTCTCTGGCGGCACTCCCTGTCTGTGGCTCATCTCTGTTGTCAGT 2525

Qy 606 rArgPheAspGlyIleGlnGlyLeuLeuGlyGluThrIleThrValArgThrLeuHi 626
Db 2526 CGCGCTTGCCTGCTGCTCTGAGGAACCTGACAGACAGGAGCTTTCAGTTGAG-----CA 2579

Qy 626 sGlnGluLeuMetThrPheAlaGluGluAlaThrPheGluPheArgLeuPheThrGlyGl 646
Db 2580 CCCACGCGCTTATCC---CCACATCTCTAGCCCGACACTGGCCCTGAGCAGCGCTC 2636

Qy 646 uValIleLysLeuPheGlnAspGlyThrPheSerIle 658
Db 2637 GTACCCCTCCGAGCTGCAGATGGCACTTTCTCTGTG 2673

RESULT 14
CNS03HC2/c
LOCUS
DEFINITION
Tetraodon nigroviridis genome survey sequence T7 end of clone
026D21 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION
AL244091
VERSION
AL244091.1 GI:7965103
KEYWORDS
GSS; genome survey sequence.
SOURCE
Tetraodon nigroviridis
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Tetraodon.
1
REFERENCE
Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bertot,A., Fzames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
```

```
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
Fzames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F.,
Saurin,W., Bertot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
20359837
PUBMED
REFERENCE
AUTHORS
Direct Submission
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
Location/Qualifiers
source
1..1052
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone_lib="G"
/clone_id="026D21"
/notes="Genoscope sequence ID : COBG026CB11LP1-end : T7"
ORIGIN
Alignment Scores:
Pred. No.: 2.64 Length: 1052
Score: 110.00 Matches: 54
Percent Similarity: 36.36% Conservative: 18
Best Local Similarity: 27.27% Mismatches: 86
Query Match: 3.12% Indels: 40
DB: 29 Gaps: 9
US-10-691-383-2 (1-676) x CNS03HC2 (1-1052)
Qy 71 IleSerAlaSerLysIleLeuGlyLysIleMetAlaArgValArgIle-----Ala 87
Db 1046 GTCCCCACGTCGAGGCGGCGGAGGCCATCGCGAGCAGAGGTCACAGCATCGAR 987

Qy 88 ThrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPheAspGluValThrAlaSer 107
Db 986 ACCACCGCGCGCACGCTCCMAGAAGCTCATACGCTTGGCGTAGAAGCGGTGTCGCGCAT 927

Qy 108 Gly-----ValPheProGluGlu 113
Db 926 GGCCATGTAGCGGTGACGTGATGAGAGCATGAGGCGAGCGTGGAAACAGAGAG 867

Qy 114 HisLysHisThrGlyGluGlyArgHisLeuGlnThrCysThrAsnSerAspAlaLeu 133
Db 866 CACCGCAT-----GAGGCCACACCTTGCA-----GCTCAGGACGCTGTA 825

Qy 134 AspPro-----ThrAlaProAsnArgAspAsnValAlaPheAlaSerArg 149
Db 824 GTTCCCGAGCGGAGCGGCTTCCGTACCGCACACAGGACGAGGAGAGCGAGCGMCCGA 765

Qy 150 ArgAspAlaAlaArgArgGluArgAspGlyThr-----GlyThrValCysGlnIle 166
Db 764 GCGAGCGCTGTCGCCAGGACAGCTCCAGCAGGAGTAGTAGGCGGCTTGTGAGCGCG 705

Qy 167 ThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSerLeuProHisAspGluLeu 186
Db 704 CGGTCTCTGAGGACGAGCGGACACC-----AGCAGGTTGCCACCCAGTGCACG 654

Qy 187 GlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCysIleLeuAsnGlyAspPhe 206
Db 653 CAGATGATGATCCCGAGCAGCAGCTTACGCTTGGCGGAGCGCGCGCGG----- 600
```

Qy	173	LeuAlaThrMetPheHisIysSerLeuProHisAspGluLeuGlyGlnValThrAlaAsp	192
Db	431	TTTGCTAACATG-----CCAAAGACATCTACCAAGATTATATCTCAAGAC	387
Qy	193	AspPheAlaIleLeuGluAsp-----CysIleLeuAsnGlyAspPheSerIle	208
Db	386	TCITTCAGTATACAAGATGGAACTCTCAGTATTGTGATTCATCAGCA---TTTATTTC	330
Qy	209	CysGluAspValProAlaGlyAspProAlaGlyArgLeuValAsnProThrAlaAlaPhe	228
Db	329	ACAAAAGACCAAGCCTGTC-----ACAGCCCAACATG	300
Qy	229	AlaIleAspIleSerGlyProAlaPheSerAlaThrThr-----Ile	242
Db	299	TATTTTGACAGTGGGAACCTGCCCCCAGCAGCACATCACAGCAGGCAGACTCTCAGTCA	240
Qy	243	ProProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyTrp	262
Db	239	CCTCCTTGAGCGCTTCACCATCACAGACATTTCTTTGAGTCAGTGTAGTGTAG-----	189
Qy	263	MetAlaLeuAlaArgAspValProPheMetGlnTyTGlyThrAspGluIleThrThr	282
Db	188	-----AACGAGTATTTTATTGTAATAAATTAACAGCAACT	156
Qy	283	AlaAlaAlaAsnLeuAlaGlyMetGlyPheProAsnLeuAspAlaValSerIleGly	302
Db	155	ATCTGGAAGAACCTTTCTTAAT-----CCAGAGATGACTTCT-----GGA	117
Qy	303	SerAsp	304
Db	116	TCTGAT	111

Search completed: September 19, 2004, 09:22:39  
Job time : 5464.35 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2004, 03:33:16 ; Search time 40.9931 Seconds  
(without alignments)  
1586.253 Million cell updates/sec

Title: US-10-691-383-2  
Perfect score: 3528  
Sequence: 1 MLCHAAADTRGSPMPDTGVL.....SIDGDMCSGLVYTGVAQCOA 676  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	127.5	3.6	650	2 H84356	long-chain fatty-a
2	120	3.4	1742	2 S76110	hypothetical prote
3	118.5	3.4	2254	2 D86215	protein T6D22.14
4	116.5	3.3	145	2 F75447	vanadium chlorop
5	115	3.3	750	2 A84315	malate dehydrogen
6	113.5	3.2	912	1 IJCHCN	N-cadherin precurs
7	113.5	3.2	3472	2 T31308	hypothetical 367K
8	113	3.2	1374	2 AB3259	extracellular seri
9	112.5	3.2	7576	2 T17428	FK506 polyketide s
10	112	3.2	1019	2 A83613	conserved hypochet
11	111	3.1	876	2 E96674	hypothetical prote
12	110.5	3.1	439	2 B44315	cartilage oligomer
13	110.5	3.1	504	2 AF1789	glucuronate kinase h
14	110.5	3.1	794	2 C87437	paniclein-binding
15	109.5	3.1	1965	2 S75200	fat protein - Syne
16	108.5	3.1	165	2 C70888	hypothetical prote
17	107.5	3.0	646	2 G83006	esterase EstA PA51
18	107.5	3.0	797	2 A96232	succinoglycan bios
19	107.5	3.0	797	2 AD3054	hypothetical prote
20	107.5	3.0	3624	2 AD0835	large repetitive p
21	107	3.0	2273	2 T09083	hemagglutinin/hemo
22	107	3.0	3972	2 S75251	hypothetical prote
23	107	3.0	5188	2 B85547	probable RTX famil
24	106.5	3.0	947	2 H86933	probable DNA topoi
25	106.5	3.0	1389	2 I50090	carboxypeptidase g
26	106	3.0	426	2 H69127	histidinol dehydro
27	106	3.0	1121	2 T06065	hypothetical prote
28	106	3.0	1122	2 T14180	exit protein - Myc,
29	106	3.0	3705	2 AD0123	probable autotrans

ALIGNMENTS

RESULT 1

H84356

long-chain fatty-acid-CoA ligase [imported] - Halobacterium sp. NRC-1

C;Species: Halobacterium sp. NRC-1

C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 24-May-2001

C;Accession: H84356

R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S

; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablon,

Jung, K.H.; Alam, M.; Freitas, T.

Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li,

A;Title: Genome sequence of Halobacterium species NRC-1.

A;Reference number: A84160; MUID:20504483; PMID:11016950

A;Accession: H84356

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-650 <STO>

A;Cross-references: GB:AE004437; NID:gl0581487; PIDN:AAG20220.1; GSPDB:GN00138

C;Genetics: lfl2

C;Superfamily: Synechocystis long-chain-fatty-acid-CoA ligase; acetate-CoA ligase homolog

Query Match	3.6%	Score 127.5;	DB 2;	Length 650;
Best Local Similarity	22.6%;	Pred. No. 0.14;		
Matches	97;	Conservative	48;	Mismatches 176;
				Indels 109;
				Gaps 22;
QY	10	RGSPMDTGVLRLTSEORAKG	-----RRQLEGEKSLGFHP-SETPYIKYLEGSETWK	62
DB	237	RFGRPRDRGDAPAVTSDSRHLSFLPLAHVLERLSGH	----FLPLASGAQVCYAESPDTLR	292
QY	63	K- - -VKLPDTDGISAKILGIMARVRIATLAVVLAAPCLAFDEVTASGVFPESHKHTGE	119	
DB	293	EDGLPEPTSATSPRVYEKLYDTIREQA	-----DSPVSERVF-EWATRVGR	339
QY	120	GRHLQTCNSDDALDTPAPNR-RDNVAFASRRDAARRRD	----GTCTVCOITNGETDLA	174
DB	340	AAH- - -TTDDPGIGLRAAHALADTLVFSIRDALGGEIEFFISGGSL	-----SPELC	389
QY	175	TMFH-KSLPHDELQGVTTADDFAILEDCLINGDSFICEDVPAGDPAGRLVNPPTAAFAIDIS	233	
DB	390	ALPHGMGLP	-----ILEG-YGLTETSPV	425
QY	234	GPFAFSATTIPPVPTLSPE	-----LAAQLAEILYMLALARDVPFMQYGTDEI	279
DB	426	GPVPTDTLTVDTETIASPEQRCDGNAGELLARGPQVFDGYW	-----GLPDA	473
QY	280	TTTAAANLAGMGFPNLDVAYSIGSDGTV	---DPFSQLFR-ATFVGVGTGTPVSVOLLVNSF	335
DB	474	TDAAFVTRCKEWFRTGDDVVELRPGDGYVRFLERAKQLLTSTGKNVAPGPPIEDAFASPL	533	
QY	336	TIDAITVEPKQETFAFDPLNVMVDFD	---EWINIQNGPFPAGPELDELFIARNDLAR	392

Db 534 VAQAMVVGDKGKFSAIL--VPNFDVSEWAASQEIAPL-----DDRDAICRDERVRAR 585  
Qy 393 V-SFVDNINT 401  
Db 586 IQSAVDDVNT 595  
RESULT 2  
S76110  
hypothetical protein - *Synechocystis* sp. (strain PCC 6803)  
C;Species: *Synechocystis* sp.  
A;Variety: PCC 6803  
C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
C;Accession: S76110  
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda DNA Res. 3, 109-136, 1996  
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp.  
A;Reference number: S74322; MUID:97061201; PMID:8905231  
A;Accession: S76110  
A;Status: nucleic acid sequence. not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-1742 <KAN>  
A;Cross-references: EMBL:D63999; GB:AB001339; NID:g1001396; PIDN:BAA10088.1; PID:d101073  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C;Genetics:  
A;Start codon: GTG  
Query Match 3.4%; Score 120; DB 2; Length 1742;  
Best Local Similarity 19.4%; Pred. No. 2.5;  
Matches 153; Conservative 77; Mismatches 242; Indels 316; Gaps 39;  
Qy 107 SGVFPBKHKTGEGRHLOCTC-----NSDDA-----LDPTAPN--RRDNVAFASRRDAARRE 156  
Db 684 NGVVTGNPANTGDCGTITATDNAGTNDVQGEIDTTPSGTPNNGNSPF-SPLDITDPT 742  
Qy 157 RDGTGTVCQITNGETDLATMFHKSPLHDELG---QVTADDEFAILED----- 199  
Db 743 DTGADDLGS--SDGNPELTFTGEGCLEIELGSGNSVDDPAYEVLETPPGVNPDPSTYTI 801  
Qy 200 CILNGDFSICEDVPAGD-----PAGR-----LVNPTAAFAIDISGPA 236  
Db 802 KLIDADSPDSDPFGDFGNGANTGDGIYITATDNAGTNDVQGEID----- 856  
Qy 237 FSATTTPPVTLSPPELAAQLAELLYMALARDVPFMQYGTDEITTTAAANLAGMGPPNL 296  
Db 857 ---TTSPTGTPNNGNS-----PFSPLDITDATDTGADLLSGNGNPFL 895  
Qy 297 -----DAVSGSDGTVDPP 310  
Db 896 TFGESGLEIELLPGDGNPVDPAYEVETPGANPGDPSTYTIKLVADPDPDSDPP 955  
Qy 311 SQLFRATFVG--VETGPFVSQLVNSPTIDA-----ITVEPKQET 348  
Db 956 GDFPVGVTGNPANTGD-----GTYITATDNAGTNDVQGEIKTSSLSINDRGT 1008  
Qy 349 FA-PDLNY-----MVDFDWLNQ-----NGPAPGPEE-----LDEE 380  
Db 1009 VWIGDLGFGSINVDLANDQGLGIQITNLQPLKGTAVLRDNGTNNFSDFLTTHIEE 1068  
Qy 381 LRFRNARDLARVSFVNINTEAVRGSLSL-----ILLELGAFSRP----- 419  
Db 1069 LLDURSATGVRIDISDFSSASTNNSVGQYVQLDHFETILDASGRLLTPADLEYRNA 1128  
Qy 420 -----GINGPFTSDRQAGFVNFSGTSHYFRILIGAELAQRA 455  
Db 1129 LASAIPGLQISKONPVQSIELPGGIVAPFLTSDQNGQVN-----HFFALIDA----- 1177  
Qy 456 SCYQKQVH--RPARPALGGT--LHNTIAGDLAD---FDISLENDEL-----LKR 502  
Db 1178 ---NPGQVDYVREA-----NGTYSFEDSAFGDRDNLRAVNIILPGQFVPPAIGLTAN 1229

Qy 503 AEINAAQNPNNEVTYLLPQ-----AIQVSGTPHSPYSGSHATONGAFATVLKALIGL 554  
Db 1230 AGVFTANGLVDQFTYITDVNGDNRNRTASVAIDNQHLKF-----TLLGV 1273  
Qy 555 DRGEGCFPNPV--FPSDDGLELNFEGACLTYSGEINKLANVA--FGRQMLGIHYRFDG 610  
Db 1274 D--GD-FKNEVAIFRVDD-----RQGTINGIAPGQAGVLEAALSNSQVIFTG 1317  
Qy 611 IQG--LLLGETIIVRTLHQELMTFAEANTEPEPLFTGEVILKFDQGTFFSDGDMCSGLVY 668  
Db 1318 ISGQTQLFGESLT-RTMEE---FKVSDSFGFLIIONNSLDM-----VLRQAAGAIS 1365  
Qy 669 TGVAQCQA 676  
Db 1366 TPVFFSQA 1373  
RESULT 3  
D86215  
protein T6D22.14 [imported] - *Arabidopsis thaliana*  
C;Species: *Arabidopsis thaliana* (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C;Accession: D86215  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.P.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.; ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: D86215  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-2254 <STO>  
A;Cross-references: GB:AE005172; NID:g8778840; PIDN:AAF9839.1; GSPDB:GN00141  
C;Genetics:  
A;Gene: T6D22.14  
A;Map position: 1  
Query Match 3.4%; Score 118.5; DB 2; Length 2254;  
Best Local Similarity 20.7%; Pred. No. 4.8;  
Matches 107; Conservative 68; Mismatches 174; Indels 167; Gaps 26;  
Qy 134 DPTAPNRDNVAFASRRDAARRE-----RDGTGTVCQITNGETDLATMFHKSPLH 183  
Db 1723 DPAANTESENLEAIEPQASAGSETVETTDFAASHQGDQVTCPLSSPTG---NQPAPE 1777  
Qy 184 DEL-GQ--VTADDFAILEDCLNGDFPSCIEDVPAGD-----PAGELVNPTAAFAIDIS 233  
Db 1778 ANTEGONINTASPHVAGPDVAVESGDYAVIDQBTMGQAQDACSLEPSG-VGTQSDLGANIE 1836  
Qy 234 GPAFSATTIPPVTLSPPELAA-----OLAEIYMWALARDVPFMQYGTDEITTTAAAN 286  
Db 1837 GQ--NTTVAQLPFGSDAVVTCGSPVSDCAQ-----DASPMPLSSPGNHPTD---AVN 1886  
Qy 287 LAGMGPPNLDVAVSI-----GSDG---TVDPFSQLFRATFV-----QVETGPFVFSOL 330  
Db 1887 IEG-----LDNTSVAEPHISGSDACEMEISEPGQOVERSTFANLPHFEGVHSGAVTAL 1940  
Qy 331 ----LVNSFTIDATVPEPKQETFPADLNYMVDPEWLNQNGGPPAGPEELDEELFIRNA 387  
Db 1941 VPSLLANGT-EQIAVQP-----VPQIPFPVFNDFPLH-----ELEKURRE 1979  
Qy 388 RDLARVSFVDNINTEAVRGSLSLLELGAFSRPGINGPFFIDSDRQAGFVNFSGTSHYFRILG 447  
Db 1980 SENSKTPEKVS-----FHYPKSIL 2001  
Qy 448 AAELAQRASCYQKQVHRFARPEALGGTTLHNTIAGDLADFDI-----SLEEN----- 495









A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen  
A;Reference number: A82950; MUID:20437337; PMID:10984043  
A;Accession: A83613  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1019 <STO>  
A;Cross-references: GB:AE004464; GB:AE004091; NID:g9946099; PIDN:AAG03651.1; GSPDB:GN0011  
A;Experimental source: strain PA01  
C;Genetics:  
A;Gene: PA0262

```
Query Match      3.2%; Score 112; DB 2; Length 1019;
Best Local Similarity 20.3%; Pred. No. 4.5;
Matches 140; Conservative 81; Mismatches 255; Indels 212; Gaps 35;

Qy 52 IKYLEG-----SETKVKVLPDGTGSAKILGKIMAR-----VRIATALAVL 94
Db 436 VDVLGDCDQPIVTRTYRATNRPYALPDHKLSTIKSKEYKGSRAHELRIDDTTAQIS 495
Qy 95 AAPCLAFDEVTASGVF-----PEEHKHTGEGRHLOTCTNSDDALDPTAPNRDNVAF 146
Db 496 AA---LMSDHGASALHGLYTHPRPEGKPRGEGFELT-----DEHGVRAAKGLL 544
Qy 147 AS-----RRDAARRRDGTGTCQITNGETDLATMF-----HKSLPHDELGVQVTDFFA 195
Db 545 LSTEEQLRAGAGHLDR---GVVVQVLEALELARELDYAGEHQGVGHDAAPQOT----- 596
Qy 196 ILEDCTUNGDFSCEDV---PAGDPAGRLVNPTAFADISGPAPSAATTIPVPVTLSPPEL 253
Db 597 -LQEAV-----RDLGHGANDSGSKNGKPAIA--LSGPAGIAAATPASLTLAAGEH 645
Qy 254 AAQLAELVWALARDVPFQYGTDEITTTAAANLAGMGFPNLDVSGISGDTVPDSOL 313
Db 646 VDSVARONQVTAQGVVINGSD-----IGLFAQGG-----EL 679
Qy 314 FRATFVGVERGPFVSQQLVNSFTIDAIVBPQK--ETFPADLNVMVDFDEWLNIQNGGPP 371
Db 680 RQIT---HQPMLLOAQKND---IRLEAKQSVESASQOHVLTAKHEHITLMCGG-- 728
Qy 372 AGPEELDEELRFRNARDLARVSVDNINTEAY---RGSLLILELGAFSRPGINGPIDS 428
Db 729 -----AYLTGKGNII--ELG-----MPG----- 744
Qy 429 DRQAGFVNGFTSHYFRILGAEALQAASQYQKQVHRFARPEALGG-----TLH 477
Db 745 ----NFVVKAAKISH---VGAASLEAELPQFVETQRRFVLKQDGTAMPNVPYITMA 798
Qy 478 N--TIAGDLADFDISLENDELKRVAEINAAQNPNNEVYL---LPQAIQVGSPT-HP 531
Db 799 NGEVIEGVTDAEGATQLLQXDAMNIAKVDKMTKSPASAVAGIAAAVGAAVAKLGGP 858
Qy 532 SYPSGHATQNGAFATVULKALIGDRGCECPNVPFSPDDGLELINF-EGACLTVEGINK 590
Db 859 DAEGRALSEGE-----ISLAG-----VFGDSIDYSTVRLRDEDEYVPMQGDYV 903
Qy 591 LAVN--VAFCRQMLGI-HYRFDGI--QGLLL-----GSTIVRTLHOSLMTFAE 634
Db 904 MAPNGHYFGEELRGVADWSLESQROGLFIHETHVWQHGVNVLVGVAYQARQFLL 963
Qy 635 EATFEPFLFTGEVILFQDGTFFSIDGDM 662
Db 964 GDQVAYRLFGKTLK---DYNIEQQGDI 988
```

RESULT 11  
E96674  
hypothetical protein F16G16.7 [imported] - *Arabidopsis thaliana*  
C;Species: *Arabidopsis thaliana* (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
A;Accession: E96674  
R;Theoretical protein F16G16.7 [imported] - *Arabidopsis thaliana*  
C;Species: *Arabidopsis thaliana* (mouse-ear cress)  
A;Title: COMP (cartilage oligomeric matrix protein) is structurally related to the thrombospondin type 1 motif (TSP1) domain of the human protein p16<sup>cas</sup>  
A;Reference number: A44315; MUID:93054522; PMID:1429587  
A;Accession: B44315  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-439 <OLD>  
A;Cross-references: EMBL:X74326; NID:g396625; PIDN:CAA52374.1; PID:g396626

Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, B.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzborg, S.B.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.; ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: E96674  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-876 <STO>  
A;Cross-references: GB:AE005173; NID:g6227005; PIDN:AAF06041.1; GSPDB:GN00141  
C;Genetics:  
A;Gene: F16G16.7  
A;Map position: 1

```
Query Match      3.1%; Score 111; DB 2; Length 876;
Best Local Similarity 20.3%; Pred. No. 4.2;
Matches 105; Conservative 68; Mismatches 184; Indels 160; Gaps 23;

Qy 221 LVNPTAFAIDISGPAPSAATTIPVTLSP-----ELAAQLAELVW----- 263
Db 59 LCNQLSPFASTTNG--LSATKNAEIPVGNPSPEERNLNETSALAAMEMMKRGLGLSE 116
Qy 264 -----ALARDVPFQYGTDEITTTAAANLAGMGFPNLDVSGISGDTVPDSOLFRAT 317
Db 117 IQDLSDIVERAVSGQLTVRELCTVRSLTATSTQKLRKAAI--SDNRVTPLVLDILO-- 173
Qy 318 FVGVEGTGPFVSQQLVNSFTIDAIVBPQKQETFPADLNVMVDFDEWLNIQNGGPPAGPEEL 377
Db 174 --GCDFKDTLQKKI--SFCIDC-----NMTMILD-----RA 200
Qy 378 DEELRFRNARDLARVSVDNINTEAYRGSLLILELGAFSRPGINGPIDSQAGVNF 437
Db 201 SEDLEIRSR-----RRNMENLDSLLKXIKTIFLAGINKPLITQRRSRMCVAI 251
Qy 438 GTSHYFRILGAEAL--AORASCVQKQVHRFARP--EALGCTLHN-----T 479
Db 252 RATHSKLLPGVVLSVSSRATC-----FIEPKAV--ELNNMEVRHANSEKAEEMA 301
Qy 480 IAGDLADFDISLENDELKRVAEINAA-----ONPNNEVYLFPQAIQV 525
Db 302 ILSILTSEVVMQAQREILHLDRILELDIAPARASHANWINGVYPNVTSEHTKTPGLAVDI 361
Qy 526 GSPTSHSPGHATQNGAFATVULKALIGDRGCECPNVP--FPDGDGLELINFEG----- 579
Db 362 DSAQHP-----LLGSLVLSGPNGGDIFPVPVDIKVESAKVVVISGPNVTGG 407
Qy 580 --ACLTVEGINKLVNVAFGQMLGIHYR-----FDGIQGLLGLGETITVRTLHQE 628
Db 408 KTAALLKTLGLLSMSKS-----GMVLPKNCPLPWFED---LILADIGDQPSLEQS 455
Qy 629 LMTFAEATFEPFLFTGEVILFQDGTFFSIDGDMCSG 665
Db 456 LSTFSGHIS--RI--RQILDIASENSLVLLDIBCSG 487
```

RESULT 12  
B44315  
cartilage oligomeric matrix protein - bovine (fragment)  
C;Species: *Bos primigenius taurus* (cattle)  
C;Date: 09-Dec-1993 #sequence\_revision 03-May-1994 #text\_change 23-Dec-2002  
A;Accession: B44315; S36710  
R;Oldberg, A.; Antonsson, P.; Lindblom, K.; Heinegard, D.  
J. Biol. Chem. 267, 22346-22350, 1992  
A;Title: COMP (cartilage oligomeric matrix protein) is structurally related to the thrombospondin type 1 motif (TSP1) domain of the human protein p16<sup>cas</sup>  
A;Reference number: A44315; MUID:93054522; PMID:1429587  
A;Accession: B44315  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-439 <OLD>  
A;Cross-references: EMBL:X74326; NID:g396625; PIDN:CAA52374.1; PID:g396626

Qy	59	ETWKVKLPDGTGTSASKILKIMARVRIATALAVLVAAPCLAFDEVFTASGVFPPEHKHT-	117
Db	42	ETGKAESPTIPEA-----VITSIQVMK-----DIDKTELGRFSFSAMHSL	85
Qy	118	-----GEGRHLOTCT-----NSDDALDPTAPNRONVAFASRRDAARERDOTGT-----	162
Db	86	IMVGADGELLTECITWADGRSSEALENI-----KRDNYLF-----OLYEATGTPIHPMS	134
Qy	163	-----VCOITNGETDLATMEKSLPHDELQGVTTADDEAILEDCLINGDFSIC-----EDVPAG	215
Db	135	PFKACIKWKEDEPOL-----FHR-----AEKFVDIKSYILNULFVGWVMDESLSAG	180
Qy	216	DPAGRLVNPTAAF-AIDI--SGPAFSATTIPPVPVPTLSSPELAAQLAELYNMALARDPFPM	272
Db	181	TGLYNILEHDFWEFAMEIILKLTDPYLPKVVPETVQLHG--VKKEYAEL--MGIENULPFI	236
Qy	273	QYGTDEITTTAAANLAGMGFPNLDVAVSIGSDGTVPDFSQLFRATFVGVTGPFVSQLLV	332
Db	237	VGGS-----GALANIGIQATGRNDVTITVGTSGAV-----RKLT	271
Qy	333	NSFTIDAITVEPKQETP---APDLNVMVDFDEWLNIQNGGPPA--GPEELDLELFRINA	387
Db	272	NEFOVDS-----RGRTFCYAGD--GYFIAGGA-----VNNGGKVBWGLOOFGESEIWN--	320

Db 846 SGTAKAIAENSGANOVYQVTTATDDADVSAGITF 880

Search completed: September 19, 2004, 03:59:48  
Job time : 44.9931 secs

RESULT 15

S75200  
fat protein - Synchocystis sp. (strain PCC 6803)  
N:Alternate names: protein slr2046  
C:Species: Synchocystis sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
C:Accession: S75200  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synchocystis  
s.  
A:Reference number: S74322; MUID:97061201; PMID:8905231  
A:Accession: S75200  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1965 <KAN>  
A:Cross-references: EMBL:D90903; GB:AB001339; NID:G1652127; PIDN:BAA17114.1; PID:d101784  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match		3.1%;	Score 109.5;	DB 2;	Length 1965;
Best Local Similarity		21.0%;	Pred. No. 19;		
Matches		146;	Conservative	78;	Mismatches 250; Indels 221; Gaps 36;
Qy	97	PCLAFDEVATAGVPEEHEKHITGEGRHLOTCNDSDDLDPAPNRDRNVAFASRRDAARRE	156		
Db	284	PSVSNNTV-V-GFPFGSQGVNLSTLTGTGPTDTPPNA-----PSTPDLSSASS	336		
Qy	157	RDGTGTCQITNGET-----DLATMFHKSLPDELGOVTAD--DFAILEDCLN	203		
Db	337	DSGLSSTDNITNDTPTFTNGTAEANSTVTLF--SGSGTQIGSTTANGSGNWTITASTPAD	394		
Qy	204	GDFSIDCEDVPAGDPAGRL-----VNPTAAFAIDISGPAF-----	237		
Db	395	GNYSI--TAKATDAAGNVSTASSALGITDNTTTPNLASAIEISDTALKIGDTATVTFTFS	452		
Qy	238	-----SATTIPVPTLSPPELAQAEALYMWALARDVPFMQYGTDEIT--TTAAANLA	288		
Db	453	EAVIGFTNADIIVVDGSLSSP--TSSDGGITWTATLPNAAESNSNVITLDTNIGISDLA	510		
Qy	289	GMGCF-----PNLDAVSISSDG-----TVDPF	310		
Db	511	GNNGTGTTTSVSYAVDIPPTLSIDDDDDNIVIPDPTLTYYTLTFSEIDSTVTATDDF	570		
Qy	311	SQLFRAIF-VG--VETGPFVSQLLVNSFTDAITVEPKQETFPD-----LNYMVFDEW	362		
Db	571	DNAGTAISIGTITETSSGVFVVVPTNSTIILQIPNGAVLSDMAGNNLAVFPVQDDDE	630		
Qy	363	LNIONGGPPA--GPEELDEELFRNRARDLARVFNINTEAYRGSLLLELGA--FSR	418		
Db	631	LQV-NOGFSAVIVPNASIAENTDTNPLKVADIAITDD-----GLGSNDISL	676		
Qy	419	PGINGPIDSROAGFVNFQSHYFRLIGAEALQARACYQKWQVHRFAPEALGGTLHN	478		
Db	677	SGIDAAFEVIGQSLFKAGFVLNFE-----SKAS-----YTVVNVDDDTVGST---	721		
Qy	479	TIAGDLADFDIISLENDL-----LKRVAEINAAQNPNNEVTYLLPQAIQVGSPT	529		
Db	722	---PDLTADFKLTINNLDEVAPTITSGDTAKAIAENSGA---NQIY-----QV----	764		
Qy	530	HPSYPSGHATONGAFATVTKALIGLDRGCECFNPVPFSPDDGLE-LINFEGACLTYEGEI	589		
Db	765	-----KATDDGDISA-----GITFG-----LKPDDADSFIIN-----ATTGEV	798		
Qy	589	NKLAVNVAFGRQMLGIHYRE-----DGIQGLLGETITVTRLHQELMTFAEEATFEFLP	643		
Db	799	-KLIGNPDPETQS---SYKFTVTADGV-----NPATEQLVNLAIIDLDEIAP---TIT	845		
Qy	644	TGEVIKLFQDG-----TFSIDGDMCSGLVY	668		

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 19, 2004, 03:56:03 ; Search time 254.636 Seconds  
(without alignments)  
3303.314 Million cell updates/sec

Title: US-10-691-383-2\_COPY\_435\_632

Perfect score: 1039

Sequence: 1 VNFGTSHYFLRIGABLAQR.....GLLGLTITVTLHQBLMTF 198

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US10691383/runat\_17092004\_102655\_1733/app\_query.fasta\_1.1230  
-DB=N\_Geneseq\_29Jan04 -QFWT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=bloms62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10691383@cgn 1.1 653 @runat\_17092004\_102655\_1733 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_29Jan04:\*

1: Geneseqn1980s:.\*  
2: Geneseqn1990s:.\*  
3: Geneseqn2000s:.\*  
4: Geneseqn2001as:.\*  
5: Geneseqn2001bs:.\*  
6: Geneseqn2002s:.\*  
7: Geneseqn2003as:.\*  
8: Geneseqn2003bs:.\*  
9: Geneseqn2003cs:.\*  
10: Geneseqn2004s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1039	100.0	2931	3 AAA10305	AAA10305 Fucus van
2	1039	100.0	2931	6 AAD28155	Aad28155 Fucus dis
3	332.5	32.0	1797	4 AAD11050	Aad11050 Corallina
4	329.5	31.7	1794	2 AAV56020	AAV56020 Haloperox
5	322	31.0	1791	2 AAV56021	AAV56021 Haloperox
C 6	108	10.4	110000	4 AAI99682_42	Continuation (43 o
C 7	108	10.4	110000	4 AAI99683_42	Continuation (43 o
8	101.5	9.8	597	7 ACA53911	ACA53911 Prokaryot

9	101	9.7	5828	3 AAA14845	Aaa14845 DNA encod
10	99	9.5	51	3 AAA10306	Aaa10306 Curvulari
11	99	9.5	51	6 AAD28156	Aad28156 Probe #1
12	91.5	8.8	1161	4 AAh22872	Aah22872 R. ruber
13	91.5	8.8	10480	4 AAh22877	Aah22877 R. ruber
14	88.5	8.5	1191	4 ABL28263	ABL28263 Drosophil
15	88.5	8.5	2117	4 AAF81756	Aaf81756 Human mem
16	88.5	8.5	2122	4 AAD12569	Aad12569 Human pro
17	88.5	8.5	2192	9 ADE31718	Ade31718 Human 646
C 18	88	8.5	819	7 ACA42710	ACA42710 Prokaryot
19	87.5	8.4	1815	6 AAh77240	Aah77240 DNA of ha
20	87.5	8.4	1815	6 ABA03638	Aba03638 Dreschler
C 21	87.5	8.4	1815	9 ADB69915	Adb69915 C. neofor
22	87	8.4	1812	6 AAS15767	Aas15767 D. salina
23	86	8.3	579	7 ABA02880	Aba02880 Dendryphi
24	86	8.3	3058	5 ACA39594	ACA39594 Prokaryot
C 25	86	8.3	3058	5 AAS06054	Aas06054 Angiotens
C 26	86	8.3	4334	5 AAS06053	Aas06053 Angiotens
27	85.5	8.2	1830	2 AAT00803	Aat00803 Chloroper
28	85.5	8.2	2088	2 AAT03875	Aat03875 Chloroper
29	85.5	8.2	2763	6 ABA92252	Aba92252 Mouse Her
30	85.5	8.2	2781	6 ABA92253	Aba92253 Her-2/neu
31	85.5	8.2	3771	3 AAh89737	Aah89737 Mouse Her
32	85.5	8.2	3771	5 AAh42210	Aah42210 Nucleotid
33	85.5	8.2	3771	6 ABA92251	Aba92251 Mouse Her
34	85.5	8.2	6273	7 ACC44344	ACC44344 Gene enco
35	85.5	8.2	12279	6 ABQ93367	ABq93367 Human cDN
36	84.5	8.1	6197	6 ABS78947	ABs78947 E. coli C
37	84	8.1	504	5 AAh68141	Aah68141 C glutami
C 38	84	8.1	309400	5 AAh68534	Aah68534 C glutami
39	83.5	8.0	525	7 ACA28777	Aca28777 Prokaryot
40	83.5	8.0	1161	4 AAh22892	Aah22892 R. ruber
41	83.5	8.0	12127	4 AAD05497	Aad05497 Human sec
42	83.5	8.0	2228	6 AAB099407	ABq99407 Human cod
43	83.5	8.0	2317	6 ABS76414	ABs76414 cDNA enco
44	83.5	8.0	2317	6 ABV99386	ABv99386 Human NOV
45	83.5	8.0	2822	2 AAT51612	Aat51612 Curvulari

ALIGNMENTS

RESULT 1

AAA10305

ID AAA10305 standard; cDNA; 2931 BP.

XX

AC AAA10305;

XX

DT 03-JUL-2000 (first entry)

XX

XX Fucus vanadium bromoperoxidase cDNA.

DE

DE Vanadium bromoperoxidase; vanadium haloperoxidase; alga; ODA;

KW o-dianisidine oxidation; vanadium ion; halogenation; antimicrobial agent;

KW phenolic adhesive production; ss.

XX

XX Fucus sp.

OS

XX

XX Key

XX Location/Qualifiers

FT CDS

FT 228..2258

FT /\*tag= a

FT /product= "Fucus vanadium bromoperoxidase"

XX

XX WO200015771-A1.

XX

XX PD 23-MAR-2000.

XX

XX 27-AUG-1999; 99WO-US019893.

XX

XX 10-SEP-1998; 98US-00151189.

XX

XX (REGC ) UNIV CALIFORNIA.

XX

XX Vreeland V, Ng KL;

```
XX WPI; 2000-271414/23.
DR P-PSDB; AAY87469.
XX
PT Fucus vanadium bromoperoxidase polypeptide and coding sequence, useful
PT for halogenating or oxidizing compounds, e.g. as antibacterial agents.
XX
PS Claim 3; Page 36-41; 46pp; English.
XX
CC This sequence represents cDNA encoding a vanadium bromoperoxidase from
CC species of Fucus. This enzyme is 73.4 kD in mass and catalyses the
CC oxidation of o-dianisidine (ODA) when complexed with a vanadium ion. The
CC cDNA encoding the enzyme was isolated from a Fucus 2-cell embryo cDNA
CC library using hybridization probes (AA10306-AL0307) corresponding to the
CC second and third regions that are conserved between Curvularia and
CC Acophyllum vanadium haloperoxidase active sites. Fucus vanadium
CC bromoperoxidase cDNAs (one full-length, two 5'-truncated) were then
CC cloned into the pET LIC (ligation independent cloning) vector for
CC expression in Escherichia coli. The vanadium bromoperoxidase proteins
CC were expressed fused to an N-terminal thioredoxin tag which optimises
CC correct folding. The Fucus vanadium bromoperoxidase has a specific
CC activity several-fold higher than other algal vanadium haloperoxidases
CC for which at least partial sequences have been reported. Vanadium
CC haloperoxidases can be used in a method for enzymatically halogenating a
CC compound, and in a method for enzymatically oxidising a compound. They
CC can be used to halogenate various substances, including proteins. They
CC can be used to produce epoxides from alkenes, halogenated ketones from
CC alkenes, to produce alpha, gamma-halohydrins from cyclopropanes, and to
CC produce dihalogenated products from alkenes and alkenes. Vanadium
CC haloperoxidases may also be used to oxidise various compounds, making
CC them useful in signal generating systems in place of horseradish
CC peroxidase. The enzymes can also be used as enzymatic antimicrobial
CC agents and in the production of phenolic adhesives
XX
SQ Sequence 2931 BP; 694 A; 776 C; 784 G; 677 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 9,66e-113 Length: 2931
Score: 1039.00 Matches: 198
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-691-383-2_COPY_435_632 (1-198) x AAA10305 (1-2931)

Qy 1 ValAsnPheGlyThrSerHisTyrPheArgLeuLeuGlyAlaAlaGluLeuAlaGlnArg 20
Db 1530 GTCAACTTCGGCAGCTCTCACTACTTCCAGATTGATAGTGCCTGGAGCTGGCGCAGCGT 1589

Qy 21 AlaSerCysTyrGlnTyrTglnValHisArgPheAlaArgProGluAlaLeuGlyGly 40
Db 1590 GCCTCGTGTACCAAAAGTGGCAGGTGCATGATTTGCACGCCCGCCGAGGCTCTCGGGGGT 1649

Qy 41 ThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGlu 60
Db 1650 ACCCTCCACACACCATCGCGGGGATCTAGATGCAGACTTCGACATCTCCCTCTTGAA 1709

Qy 61 AsnAspGluLeuLeuLysArgValAlaGluLeuAsnAlaAlaGlnAsnProAsnAsnGlu 80
Db 1710 AATGATGAGCTCTTGAACAGTGTGGCGAGATAAATCGCGCAGAGATCCCAACACAGAG 1769

Qy 81 ValThrTyrLeuLeuProGlnAlaGlnValGlySerProThrHisProSerTyrPro 100
Db 1770 GTACACTTACCTTCTCCACAGCTATCCAGTGGATCGCAAGCCCTTCCTACCCG 1829

Qy 101 SerGlyHisAlaThrGlnAsnGlyValaPheAlaThrValLeuLysAlaLeuLeuGlyLeu 120
Db 1930 TCCGCCACGCTACCCAAATGGAGCATTTGCCACAGTCTGAAGGCCCTCATTTGGGCTA 1889

Qy 121 AspArgGlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuLeuLeu 140
Db 1890 GATCGGGAGGTGAGTGTCTTCCCTTAACCCCGTGTTCCTCAAGCGATGATGCGGCTGGAATA 1949
```

```
Qy 141 IleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsn 160
Db 1950 ATCAACTTCGAAGGGGATGCCCTTATATAGAGGAGATCAACAGACTCGGGCTCAAC 2009

Qy 161 ValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeu 180
Db 2010 GTCGCAATTTGGAGGAGGATGCTGGGATCCATCTATCGGTTCGACGGTATCAAGGCTA 2069

Qy 181 LeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGlnLeuMetThrPhe 198
Db 2070 CTTCTCGGAGAGACAATCACTGTACGAACAATTCACGAGGAGCTGATGACGTTTC 2123

RESULT 2
AAD28155
ID AAD28155 standard; cDNA; 2931 BP.
XX
AC AAD28155;
XX
DT 07-AUG-2003 (revised)
DT 01-MAY-2002 (first entry)
XX
DE Fucus distiches. vanadium haloperoxidase encoding cDNA.
XX
KW Vanadium haloperoxidase; vanadium bromoperoxidase; o-dianisidine; ODA;
KW epoxide; antimicrobial agent; phenolic adhesive; halohydrin;
KW industrial catalysis; enzyme; catalyst; ss.
XX
OS Fucus distichus.
XX
FH Key Location/Qualifiers
FT CDS 228..2258
FT /tag= a
FT /product= "vanadium haloperoxidase"
XX
FN WO200200838-A2.
XX
PD 03-JAN-2002.
XX
PF 07-JUN-2001; 2001WO-US018602.
XX
PR 19-JUN-2000; 2000US-00596794.
XX
PA (REGC ) UNIV CALIFORNIA.
XX
PI Vreeland V;
DR WPI; 2002-154651/20.
DR P-PSDB; AAE17501.
XX
PT Isolated nucleic acid for, e.g. to halogenate various substrates
PT including proteins, has polynucleotide sequence encoding vanadium
PT haloperoxidase polypeptide with catalytic helical frame.
XX
PS Claim 4; Page 49-53; 56pp; English.
XX
CC The invention relates to nucleic acid encoding vanadium haloperoxidase
CC particularly vanadium bromoperoxidase polypeptide which consists of a
CC catalytic helical frame that complexes a vanadium ion and catalyses the
CC oxidation of o-dianisidine (ODA). The invention also provides
CC recombinantly produced vanadium haloperoxidases. The polypeptides of the
CC invention can be fused to other proteins to allow quantification or
CC localisation of the linked protein. They can be used to halogenate
CC various substrates including proteins; to produce epoxides from alkenes,
CC halogenated ketones from alkenes; to produce alpha, gamma-halohydrins
CC from cyclopropanes and to produce dihalogenated products from alkenes and
CC alkenes. They can also be used in signal generating systems in place of
CC horseradish peroxidase; as a component in assays; as enzymatic
CC antimicrobial agents; and in the production of phenolic adhesives. The
CC polypeptides can be used in industrial catalysis in a variety of
CC contexts, e.g. catalyst for halogenation, oxidation, and epoxidation
CC reactions. The present sequence is Fucus distiches vanadium
CC haloperoxidase encoding cDNA. (Updated on 07-AUG-2003 to correct OS
```



```

CC field.)
XX SQ Sequence 2931 BP; 694 A; 776 C; 784 G; 677 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 9.66e-113 Length: 2931
Score: 1039.00 Matches: 198
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-691-383-2_COPY_435_632 (1-198) x AAD28155 (1-2931)
Qy 1 ValAsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArg 20
Db 1530 GTCAACTTCGGCAGCTCTCACTACTTCAGATTGATAGTGCCGCCGAGCTGGCGACGCT 1589
Qy 21 AlaSerCysTyrGlnIlysTrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGly 40
Db 1590 GCCTCGTGTACCAAAAGTGGCAGGTGCATCGATTTCGACGCCGCCGAGGCTCTCGGGGT 1649
Qy 41 ThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGlu 60
Db 1650 ACCCTCCACAAACACCATCGCGGGGATCTAGATGACAGCTTCGACATCTCCCTTCITGAA 1709
Qy 61 AsnAspGluLeuLeuIlyAsArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGlu 80
Db 1710 AATGATGAGCTCTTGAACGCTGTGGCGGAGATTAATTCGGCGCAGATCCACACAGAG 1769
Qy 81 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 100
Db 1770 GTCACTACCTTCTTCACAAAGCTATCCAAAGTGGATCGCAACGCCACCTTCTACCCG 1829
Qy 101 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuIlyAlaLeuIleGlyLeu 120
Db 1830 TCCGGCCACGCTACCCAAAATGGAGCATTTGCCACAGTTCGAAGGCCCTCATTTGGCCTA 1889
Qy 121 AspArgGlyGlyGluCysPheProAsnProValPheProSerAspGlyLeuGluLeu 140
Db 1890 GATCGGGAGGTGAGTGTCTTCCCTAACCCGCTGTCCACAGCGATGACGCGCTGGAATA 1949
Qy 141 IleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnIlyLeuAlaValAsn 160
Db 1950 ATCAACTTCGAAGGGCATGCTTACATATGAGGAGAGATCAACAAGCTCGCGGTCAAC 2009
Qy 161 ValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeu 180
Db 2010 GTCGCATTTGGAGGCGAGATGCTGGGCATCCACTATTCGGTTCGACGGTATCCAAAGGCCTA 2069
Qy 181 LeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThrPhe 198
Db 2070 CTTCCTCGGAGACAAATCACTGTACGAAACACTTCACCAAGAGCTGATGAGCTTC 2123

RESULT 3
AAD11050
ID AAD11050 standard; cDNA; 1797 BP.
XX AC AAD11050;
XX DT 24-SEP-2001 (first entry)
XX DE Corallina officinalis vanadium bromoperoxidase (cVBPO) cDNA.
XX KW Marine red algae; vanadium bromoperoxidase; VBPO; haloperoxidase;
XX KW signal generation; analytical system; anti-microbial agent;
XX KW industrial application; biological material; avidin; metabolite;
XX KW lipopolysaccharide; human chorionic gonadotropin; luteinising hormone;
XX KW vitamin; intact cell; drug; narcotic; ss.
XX OS Corallina officinalis.
XX FH Key Location/Qualifiers

CDS
FT 1..1797
FT /*tag= a
FT /product= "Corallina officinalis vanadium bromoperoxidase
FT (VBPO)"
FT primer_bind
FT complement(1..21)
FT /*tag= b
FT /bound_moiety= "PCR forward primer (AAD11054)"
FT primer_bind
FT 1150..1170
FT /*tag= c
FT /bound_moiety= "PCR reverse primer (AAD11055)"
FT misc_binding
FT complement(1420..1449)
FT /*tag= d
FT /bound_moiety= "PCR probe (AAD11053)"
FT primer_bind
FT complement(1420..1442)
FT /*tag= e
FT /bound_moiety= "PCR upstream primer (AAD11051)"
FT primer_bind
FT 1642..1664
FT /*tag= f
FT /bound_moiety= "PCR downstream primer (AAD11052)"
FT XX
FT PN WO200153494-A2.
FT XX
FT PD 26-JUL-2001.
FT XX
FT PF 24-JAN-2001; 2001WO-US002383.
FT XX
FT PR 24-JAN-2000; 2000US-0177764P.
FT XX
FT PA (REGC ) UNIV CALIFORNIA.
FT XX
FT PI Butler A, Carter JN;
FT XX
FT DR WPI; 2001-457611/49.
FT XX
FT DR P-PSDB; AAE05750.
FT XX
FT PT Novel vanadium haloperoxidase, especially vanadium bromoperoxidase,
FT obtained from Corallina officinalis for use in signal generation in
FT analytical systems.
FT XX
FT PS Claim 25; Fig 1; 68pp; English.
FT XX
CC The present sequence is Corallina officinalis (marine red algae) vanadium
CC bromoperoxidase (cVBPO) cDNA. The vanadium haloperoxidase, especially
CC vanadium bromoperoxidase (VBPO) is useful for signal generation in
CC analytical systems, and as anti-microbial agents. The VBPO is also useful
CC for industrial applications. The VBPO is also useful for detecting a wide
CC variety of chemical and biological materials such as amino acids,
CC peptides, polypeptides, proteins including enzymes, avidin, antibodies
CC and antigenic proteins, carbohydrates including monosaccharides,
CC polysaccharides and lipopolysaccharides, hormones such as human chorionic
CC gonadotropin, thyroid stimulating hormone, luteinising hormone,
CC thyroxine, follicle stimulating hormone, parathyroid hormone and growth
CC hormone, metabolites such as glucose, lactate and pyruvate
CC oligonucleotides, nucleic acids, vitamins such as B12 and biotin, intact
CC cells from various organisms including microorganisms and drugs such as
CC narcotics, therapeutic and those abused
FT SQ Sequence 1797 BP; 394 A; 480 C; 513 G; 410 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3.8e-29 Length: 1797
Score: 332.50 Matches: 82
Percent Similarity: 56.34% Conservative: 38
Best Local Similarity: 38.50% Mismatches: 72
Query Match: 32.00% Indels: 21
DB: 4 Gaps: 7

US-10-691-383-2_COPY_435_632 (1-198) x AAD11050 (1-1797)
Qy 1 ValAsnPheGlyThrSerHisTyrPheArgLeuIle---GlyAlaAlaGluLeuAlaGln 19
Db 1120 GTGAACCTTCGGAGACGCACACGCTGCTGAGTCTGTGCTGAGGTGCTACGCCGCGTGTG 1179

```





```
WP AAI99682_43 4300001 4410000
WP AAI99682_44 4400001 4411529

Alignment Scores:
Pred. No.: 5.26 Length: 110000
Score: 108.00 Matches: 46
Percent Similarity: 35.42% Conservative: 22
Best Local Similarity: 23.96% Mismatches: 52
Query Match: 10.39% Indels: 72
DB: 4 Gaps: 8

US-10-691-383-2_COPY_435_632 (1-198) x AAI99682_42 (1-110000)

QY 4 GlyThrSerHisTyr-----PheArgLeuIleGlyAla 14
Db 70277 GGAATCTCCCACTTCGCGAGCACTGCATCGTTGGCTGATCTGCGCTGCTGGCGCG 70218

QY 15 AlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArg 34
Db 70217 ATAGCCCTGCGCGCGCGCG-----CGGAATGGCTTGTG-----70182

QY 35 ProGluAlaLeuGlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPhe 54
Db 70181 ---CGCGGGCGCGCGCGCTTCGTCGCCCATGCCATCGC-----70146

QY 55 AspIleSerLeuLeuGluAsnAspGluLeuLysArgValAlaGluIleAsnAlaAala 74
Db 70145 -----GTGCTGATCAAGCGGCTGTGGCGG-----CGT 70119

QY 75 GlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySerPro 94
Db 70118 CAGCGGCGGATCATCCGCGCATC-----CGGTACACGTGGACACGCCA 70074

QY 95 ThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeu 114
Db 70073 AGTCACTGAGCTTTCGTCGGCACACGCCACCTCGACACCGCGCGCGCTGCTCATG 70014

QY 115 LysAlaLeuIleGlyLeuAspArgGlyGlyCysPheProAsnProValPheProSer 134
Db 70013 GGCAGACCCACCGGCTG-----CGGTACCGGTTGTG-----69981

QY 135 AspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIle 154
Db 69981 -----69981

QY 155 AsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPhe 174
Db 69980 -----CTGGTGGCGCGGATGCGGCTCGCGAATACTGCTGGGGTCCACTACCCAGT 69927

QY 175 AspGlyIleGlnGlyLeuLeuLeuGlyGluThrIle 186
Db 69926 GATGTGCCGTGGGTGTGTTGCTCTGGCGGCCACCGTC 69891

RESULT 7
AAI99683_42/c
Continuation (43 of 44) of AAI99683 from base 4200001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683
WP Fragment Name Begin End
WP AAI99683_00 1 110000
WP AAI99683_01 100001 210000
WP AAI99683_02 200001 310000
WP AAI99683_03 300001 410000
WP AAI99683_04 400001 510000
WP AAI99683_05 500001 610000
WP AAI99683_06 600001 710000
WP AAI99683_07 700001 810000
WP AAI99683_08 800001 910000
WP AAI99683_09 900001 1010000
WP AAI99683_10 1000001 1110000
WP AAI99683_11 1100001 1210000
WP AAI99683_12 1200001 1310000
WP AAI99683_13 1300001 1410000
WP AAI99683_14 1400001 1510000
```

```
WP AAI99683_15 1500001 1610000
WP AAI99683_16 1600001 1710000
WP AAI99683_17 1700001 1810000
WP AAI99683_18 1800001 1910000
WP AAI99683_19 1900001 2010000
WP AAI99683_20 2000001 2110000
WP AAI99683_21 2100001 2210000
WP AAI99683_22 2200001 2310000
WP AAI99683_23 2300001 2410000
WP AAI99683_24 2400001 2510000
WP AAI99683_25 2500001 2610000
WP AAI99683_26 2600001 2710000
WP AAI99683_27 2700001 2810000
WP AAI99683_28 2800001 2910000
WP AAI99683_29 2900001 3010000
WP AAI99683_30 3000001 3110000
WP AAI99683_31 3100001 3210000
WP AAI99683_32 3200001 3310000
WP AAI99683_33 3300001 3410000
WP AAI99683_34 3400001 3510000
WP AAI99683_35 3500001 3610000
WP AAI99683_36 3600001 3710000
WP AAI99683_37 3700001 3810000
WP AAI99683_38 3800001 3910000
WP AAI99683_39 3900001 4010000
WP AAI99683_40 4000001 4110000
WP AAI99683_41 4100001 4210000
WP AAI99683_42 4200001 4310000
WP AAI99683_43 4300001 4403765

Alignment Scores:
Pred. No.: 5.26 Length: 110000
Score: 108.00 Matches: 46
Percent Similarity: 35.42% Conservative: 22
Best Local Similarity: 23.96% Mismatches: 52
Query Match: 10.39% Indels: 72
DB: 4 Gaps: 8

US-10-691-383-2_COPY_435_632 (1-198) x AAI99683_42 (1-110000)

QY 4 GlyThrSerHisTyr-----PheArgLeuIleGlyAla 14
Db 62534 GGAATCTCCCACTTCGCGAGCACTGCATCGTTGGCTGATCTGCGCTGCTGGCGCG 62475

QY 15 AlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArg 34
Db 62474 ATAGCCCTGCGCGCGCGCG-----CGGAATGGCTTGTG-----62439

QY 35 ProGluAlaLeuGlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPhe 54
Db 62438 ---CGCGGGCGCGCGGCTGCTGCGCCATGCCATCGC-----62403

QY 55 AspIleSerLeuLeuGluAsnAspGluLeuLysArgValAlaGluIleAsnAlaAala 74
Db 62402 -----GTGCTGATCAAGCGGCTGTGGCGG-----CGT 62376

QY 75 GlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySerPro 94
Db 62375 CAGCGGCGGATCATCCGCGCATC-----CGGTACACGTGGACACGCCA 62331

QY 95 ThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeu 114
Db 62330 AGTCACTGAGCTTTCGTCGGCACACGCCACCTCGACACCGCGCGCGCTGCTCATG 62271

QY 115 LysAlaLeuIleGlyLeuAspArgGlyGlyCysPheProAsnProValPheProSer 134
Db 62270 GGCAGACCCACCGGCTG-----CGGTACCGGTTGTG-----62238

QY 135 AspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIle 154
Db 62238 -----62238

QY 155 AsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPhe 174
```

Db 62237 -----CTGTGCGCCGATGCGCTGCGAATACTGTGGGTGCCACTACCCAGT 62184  
Qy 175 AspGlylleGlnGlyLeuLeuGlyGluThrIle 186  
Db 62183 GATGTGCGCGGTGTGTCTGTGGGCCACCGTC 62148

RESULT 8  
ACA53911  
ID ACA53911 standard; DNA; 597 BP.  
AC ACA53911;  
AC ACA53911;  
XX  
XX  
XX 19-JUN-2003 (first entry)  
XX  
XX Prokaryotic essential gene #35568.  
XX  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
XX drug design; gene.  
XX  
XX Yersinia pestis.  
XX  
XX WO200277183-A2.  
XX  
XX 03-OCT-2002.  
XX  
XX 21-MAR-2002; 2002WO-US009107.  
XX  
XX 21-MAR-2001; 2001US-00815242.  
XX  
XX 06-SEP-2001; 2001US-00948993.  
XX  
XX 25-OCT-2001; 2001US-0342923P.  
XX  
XX 08-FEB-2002; 2002US-00072851.  
XX  
XX 06-MAR-2002; 2002US-0362699P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlseen KL, Zyskind JW;  
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
XX WPI; 2003-029926/02.  
XX  
XX P-PSDB; ABUS0041.  
XX

XX New antisense nucleic acids, useful for identifying proteins or screening  
XX for homologous nucleic acids required for cellular proliferation to  
XX isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 14; SEQ ID NO 41781; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
XX the 6213 antisense sequences given in the specification where expression  
XX of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX (1) a vector comprising a promoter operably linked to the nucleic acid  
XX encoding a polypeptide whose expression is inhibited by the antisense  
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX polypeptide or its fragment whose expression is inhibited by the  
XX antisense nucleic acid; (4) an antibody capable of specifically binding  
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX proliferation or the activity of a gene in an operon required for  
XX proliferation; (7) identifying a compound that influences the activity of  
XX the gene product or that has an activity against a biological pathway  
XX required for proliferation, or that inhibits cellular proliferation; (8)  
XX identifying a gene required for cellular proliferation or the biological  
XX pathway in which a proliferation-required gene or its gene product lies  
XX or a gene on which the test compound that inhibits proliferation of an  
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX compound's activity; (11) a culture comprising strains in which the gene  
XX product is overexpressed or underexpressed; (12) determining the extent  
XX to which each of the strains is present in a culture or collection of  
XX strains; or (13) identifying the target of a compound that inhibits the  
XX proliferation of an organism. The antisense nucleic acids are useful for  
XX identifying proteins or screening for homologous nucleic acids required  
XX for cellular proliferation to isolate candidate molecules for rational  
XX drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than S. aureus, S. typhimurium,  
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 597 BP; 115 A; 144 C; 138 G; 200 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 0.0221 Length: 597  
Score: 101.50 Matches: 39  
Percent Similarity: 42.86% Conservative: 18  
Best Local Similarity: 29.32% Mismatches: 46  
Query Match: 9.77% Indels: 30  
DB: 7 Gaps: 5

US-10-691-383-2\_COPY\_435\_632 (1-198) x ACA53911 (1-597)

Qy 77 ProAenAenGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHis 96  
Db 87 CCTGATTATGCTCATCCCTATTTTACTGATAGGGCTGTGGCTATGGGGCCAAAGATAC 146  
Qy 97 ProSerTyrProSerGlyHisAlaThrGlnAsnGly----- 108  
Db 147 -----CATGGACATGCAACGTACTGTGTGTCAGCAAGACAGCTATCGCTCTGCG 194  
Qy 109 -----AlaPhe-AlaThrValLeuLysAlaLeuIleGlyLeuAspArgGly----- 123  
Db 195 TTTTTCATGCTTTCTGCTACCTGATCGGATGCTCATCTCCTCATGATCGGCTTTGT 254  
Qy 124 -----GlyGluCysPhe-----ProAsnProValPheProSerAspAs 136  
Db 255 CGATGGCTTTGGTTATGCTTTATGAGCCATGCGCCAGACAGCTCTTTCCAGTGACCA 314  
Qy 136 pGlyLeuGluLeuIleAenPheGluGlyAlaCysLeuThrTyrGluGly----- 152  
Db 315 CGGACCGCTATTTTCTACCTTTTGCACGTGGCTTTCTTTTGGCATCACCTTTGGTCCGG 374  
Qy 153 ----GluIleAsnLysLeuAlaValAenValAlaPheGlyArgGlnMetLeuGlyIleHi 171  
Db 375 TGTGAGCTGATGTTTATGCCATTTGGATGCTGCTGATGCTATCTATCTATCTGCGTGTCA 434  
Qy 171 sTyrArgPheAspGlyIleGlnGlyLeuLeuLeuGly 183  
Db 435 TTGSCCATTCGACATGCTGGTGGTCTTCTGTAGGT 471

RESULT 9  
AAA14845  
ID AAA14845 standard; DNA; 5828 BP.  
AC AAA14845;  
XX  
XX 08-AUG-2000 (first entry)  
XX  
XX DNA encoding toluene monooxygenase proteins.  
XX  
XX Toluene monooxygenase; TomK; TomL; TomM; TomN; TomO; TomP; TomQ; toluene;  
XX ortho-cresol; 3-methylcatechol; chlorinated compound;  
XX aliphatic hydrocarbon; trichloroethylene; dichloroethylene; phenol;  
XX aromatic compound; benzene; cresol; environmental remediation; ss.  
XX  
XX Burkholderia cepacia.  
XX  
XX Key Location/Qualifiers  
XX CDS 216..446  
XX /\*tag= a  
XX /product= "TomK protein"  
XX /note= "encodes AAY84816"  
XX CDS 391..1458  
XX /\*tag= b  
XX /product= "TomL protein"  
XX /note= "encodes AAY84817"



CC are conserved between *Curvularia* and *Ascophyllium* vanadium haloperoxidase  
 CC active sites. Fucus vanadium bromoperoxidase cDNAs (one full-length, two  
 CC 5'-truncated) were then cloned into the pET LIC (ligation independent  
 CC cloning) vector for expression in *Escherichia coli*. The vanadium  
 CC bromoperoxidase proteins were expressed fused to an N-terminal  
 CC thiorodoxin tag which optimises correct folding. The Fucus vanadium  
 CC bromoperoxidase has a specific activity several-fold higher than other  
 CC algal vanadium haloperoxidases for which at least partial sequences have  
 CC been reported. Vanadium haloperoxidases can be used in a method for  
 CC enzymatically halogenating a compound, and in a method for enzymatically  
 CC oxidising a compound. They can be used to halogenate various substances,  
 CC including proteins. They can be used to produce epoxides from alkenes,  
 CC halogenated ketones from alkynes, to produce alpha, gamma-halohydrins  
 CC from cyclopropanes, and to produce dihalogenated products from alkenes  
 CC and alkynes. Vanadium haloperoxidases may also be used to oxidise various  
 CC compounds, making them useful in signal generating systems in place of  
 CC horseradish peroxidase. The enzymes can also be used as enzymatic  
 CC antimicrobial agents and in the production of phenolic adhesives. The  
 CC present sequence represents a hybridisation probe corresponding to the  
 CC second region of conservation between the active sites of *Curvularia*  
 CC *inaequalis* vanadium chloroperoxidase and *Ascophyllium* cDNA  
 CC bromoperoxidase, which was used in the isolation of cDNA encoding Fucus  
 CC vanadium bromoperoxidase. (Updated on 06-AUG-2003 to correct OS field.)  
 XX  
 SQ Sequence 51 BP; 12 A; 21 C; 9 G; 9 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 0.00143 Length: 51  
 Score: 99.00 Matches: 17  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 9.53% Indels: 0  
 DB: 3 Gaps: 0

US-10-691-383-2\_COPY\_435\_632 (1-198) x AAA10306 (1-51)

Qy 94 ProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPhe 110  
 Db 1 CCAACGCACCGCTTCGTACCGCTCTGGCCAGCTACCCAAACGAGCATTT 51

RESULT 11

AAH28156  
 ID AAD28156 standard; DNA; 51 BP.

XX AAD28156;

XX 01-MAY-2002 (first entry)

XX Probe #1 used for cloning Fucus *distichus* vanadium haloperoxidase gene.

XX Vanadium haloperoxidase; vanadium bromoperoxidase; O-dianisidine; ODA;  
 KW epoxide; antimicrobial agent; phenolic adhesive; halohydrin;  
 KW industrial catalysis; catalyst; enzyme; probe; ss.

XX *Ascophyllium* sp.

OS *Curvularia inaequalis*.

XX WO200200838-A2.

XX 03-JAN-2002.

XX 07-JUN-2001; 2001WO-US018602.

XX 19-JUN-2000; 2000US-00596794.

XX (REGC ) UNIV CALIFORNIA.

XX Vreeland V;

XX WPI; 2002-154651/20.

XX Isolated nucleic acid for, e.g. to halogenate various substrates  
 PT including proteins, has polynucleotide sequence encoding vanadium

PT haloperoxidase polypeptide with catalytic helical frame.

XX Example 1; Page 28; 56pp; English.

XX The invention relates to nucleic acid encoding vanadium haloperoxidase  
 CC particularly vanadium bromoperoxidase polypeptide which consists of a  
 CC catalytic helical frame that complexes a vanadium ion and catalyses the  
 CC oxidation of O-dianisidine (ODA). The invention also provides  
 CC recombinantly produced vanadium haloperoxidases. The polypeptides of the  
 CC invention can be fused to other proteins to allow quantification or  
 CC localisation of the linked protein. They can be used to halogenate  
 CC various substrates including proteins; to produce epoxides from alkenes,  
 CC halogenated ketones from alkynes; to produce alpha, gamma-halohydrins  
 CC from cyclopropanes and to produce dihalogenated products from alkenes and  
 CC alkynes. They can also be used in signal generating systems in place of  
 CC horseradish peroxidase; as a component in assays; as enzymatic  
 CC antimicrobial agents; and in the production of phenolic adhesives. The  
 CC polypeptides can be used in industrial catalysis in a variety of  
 CC contexts, e.g. catalyst for halogenation, oxidation, and epoxidation  
 CC reactions. The present sequence is a probe used for cloning Fucus  
 CC *distichus* vanadium bromoperoxidase gene. This probe is designed based on  
 CC the second conserved regions between *Ascophyllium* sp. and *Curvularia*  
 CC *inaequalis* vanadium peroxidase active site  
 XX

SQ Sequence 51 BP; 12 A; 21 C; 9 G; 9 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 0.00143 Length: 51  
 Score: 99.00 Matches: 17  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 9.53% Indels: 0  
 DB: 6 Gaps: 0

US-10-691-383-2\_COPY\_435\_632 (1-198) x AAD28156 (1-51)

Qy 94 ProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPhe 110  
 Db 1 CCAACGCACCGCTTCGTACCGCTCTGGCCAGCTACCCAAACGAGCATTT 51

RESULT 12

AAH22872  
 ID AAH22872 standard; DNA; 1161 BP.

XX AAH22872;

XX 17-SEP-2001 (first entry)

XX R. ruber cddB gene ORF encoding lauryl lactone esterase enzyme.

XX Cyclododecanone; degradation; lauryl lactone esterase; cdda; cddb; cddx;  
 KW dodecanol diacid; cyclododecanone monooxygenase; cyclic ketone; cddy;  
 KW 12-hydroxylauric acid dehydrogenase; 12-oxo lauric acid dehydrogenase;  
 KW macrolactone; microbiological; cddc; cddd; ds.

XX *Rhodococcus ruber*.

XX Key Location/Qualifiers

FT CDS 1..1161

FT /\*tag= a

FT /gene= "cddb"

FT /product= " lauryl lactone esterase"

XX WO200142436-A2.

XX 14-JUN-2001.

XX 08-DEC-2000; 2000WO-US033426.

XX 10-DEC-1999; 99US-0170214P.

XX (DUFO ) DU PONT DE NEMOURS & CO E I.

```
PI Chen MW, Cheng Q, Gibson KJ, Kostichka KN, Thomas SM;
XX
DR WPI; 2001-451630/48.
DR P-PSDB; AAB85325.
XX
PT Isolated nucleic acids encoding dodecanoic diacid synthesizing enzyme,
PT cyclododecanone monooxygenase for bioproduction of dodecanoic diacid from
PT cyclododecanone.
XX
PS Claim 8; Page 53-54; 78pp; English.
XX
CC The invention relates to genes from Rhodococcus ruber involved in
CC cyclododecanone degradation pathway. A 10 kb cluster from R. ruber was
CC isolated that comprises genes (designated cddA, cddB, cddX, cddY, cddC
CC and cddD) encoding enzymes such as dodecanoic diacid synthesizing enzyme,
CC cyclododecanone monooxygenase, lauryl lactone esterase, 12-hydroxylauric
CC acid dehydrogenase, 12-oxo lauric acid dehydrogenase, 12-hydroxylauric
CC acid when contacted with cyclododecanone, lauryl lactone when contacted
CC with cyclododecanone, 12-hydroxy lauric acid when contacted with lauryl
CC lactone, 12-oxo lauric acid when contacted with 12-hydroxy lauric acid
CC and dodecanoic diacid when contacted with 12-oxo lauric acid. They are
CC also useful for the production of hydroxy acids when contacted with
CC cyclic ketones with 6 to eight carbon atoms like C6, C10, C11, C12, C13
CC and C15 cyclic ketones, cyclohexanone, cyclodecanone, cycloundecanone,
CC cyclo dodecanone, cyclotridecanone and cyclopentadecanone. The host cells
CC are also used for the production of macro lactones with at least 10 carbon
CC atoms when contacted with a cyclic ketone. The nucleic acids are useful
CC for the bioproduction of dodecanoic diacid from cyclododecanone by
CC microbiological means. The present sequence represents the nucleotide
CC sequence of ORF2 (cddB gene) encoding a lauryl lactone esterase enzyme
CC isolated from a 10kb nucleic acid fragment from R. ruber SC1
XX
SQ Sequence 1161 BP; 227 A; 409 C; 353 G; 172 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0.856 Length: 1161
Score: 91.50 Matches: 50
Percent Similarity: 39.66% Conservatives: 21
Best Local Similarity: 27.93% Mismatches: 74
Query Match: 8.81% Indels: 34
DB: Gaps: 7

US-10-691-383-2_COPY_435_632 (1-198) x AAH22872 (1-1161)
QY 14 AlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAla 33
DB 453 GCGCGCGGGTTCCGGCCCGCGGTACCAACGCGCAACACACAGGCGCCACCTCATCGGAGAAAT 512
QY 34 ArgPro-----GluAlaLeuGlyGlyThrLeuHisAsnThr-IleAlaGlyAs 49
DB 513 CGTCCCGCGGATACCGCGCGCGACGGTTCAAGGAGTTCGTCCACACACATCGCGCGTCC 572
QY 49 pLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuLeuArgValAl 69
DB 573 GCTCGTCCGCACTTCAGATCGCGCGCGCGGAGAACGAT-----TGGGGCGCGTACCGC 626
QY 69 aGluLeuAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaI 89
DB 627 CGAGATCGTCCGACACCGCCCGCTTCGACATCGATCTCGCGCACTCGATCCGACGCGT 686
QY 89 eGlnValGlySerProThrHisPro-----SerTyrProSer-- 101
DB 687 CATGGTAGACGATGATACGGGACCGGTAGCGGACCGGCGGCAACTACTCCCGATG 746
QY 102 -----GlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAl 116
DB 747 GCGGACGCGGACATGGAGCCCTCAACGGTCAACGCGACGCGACGCTCGTCTCGGAT 806
QY 116 aLeuLeuGlyLeuAspArgGlyGlyGluCysPheProAsnProValPheProSerAspAs 136
DB 807 CCTCTGACCATCATCGTGGCGGTGAG-----AGCAA 839

Chen MW, Cheng Q, Gibson KJ, Kostichka KN, Thomas SM;
XX
DR WPI; 2001-451630/48.
DR P-PSDB; AAB85325.
XX
PT Isolated nucleic acids encoding dodecanoic diacid synthesizing enzyme,
PT cyclododecanone monooxygenase for bioproduction of dodecanoic diacid from
PT cyclododecanone.
XX
PS Claim 8; Page 53-54; 78pp; English.
XX
CC The invention relates to genes from Rhodococcus ruber involved in
CC cyclododecanone degradation pathway. A 10 kb cluster from R. ruber was
CC isolated that comprises genes (designated cddA, cddB, cddX, cddY, cddC
CC and cddD) encoding enzymes such as dodecanoic diacid synthesizing enzyme,
CC cyclododecanone monooxygenase, lauryl lactone esterase, 12-hydroxylauric
CC acid dehydrogenase, 12-oxo lauric acid dehydrogenase, 12-hydroxylauric
CC acid when contacted with cyclododecanone, lauryl lactone when contacted
CC with cyclododecanone, 12-hydroxy lauric acid when contacted with lauryl
CC lactone, 12-oxo lauric acid when contacted with 12-hydroxy lauric acid
CC and dodecanoic diacid when contacted with 12-oxo lauric acid. They are
CC also useful for the production of hydroxy acids when contacted with
CC cyclic ketones with 6 to eight carbon atoms like C6, C10, C11, C12, C13
CC and C15 cyclic ketones, cyclohexanone, cyclodecanone, cycloundecanone,
CC cyclo dodecanone, cyclotridecanone and cyclopentadecanone. The host cells
CC are also used for the production of macro lactones with at least 10 carbon
CC atoms when contacted with a cyclic ketone. The nucleic acids are useful
CC for the bioproduction of dodecanoic diacid from cyclododecanone by
CC microbiological means. The present sequence represents the nucleotide
CC sequence of ORF2 (cddB gene) encoding a lauryl lactone esterase enzyme
CC isolated from a 10kb nucleic acid fragment from R. ruber SC1
XX
SQ Sequence 1161 BP; 227 A; 409 C; 353 G; 172 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0.856 Length: 1161
Score: 91.50 Matches: 50
Percent Similarity: 39.66% Conservatives: 21
Best Local Similarity: 27.93% Mismatches: 74
Query Match: 8.81% Indels: 34
DB: Gaps: 7

US-10-691-383-2_COPY_435_632 (1-198) x AAH22872 (1-1161)
QY 14 AlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAla 33
DB 453 GCGCGCGGGTTCCGGCCCGCGGTACCAACGCGCAACACACAGGCGCCACCTCATCGGAGAAAT 512
QY 34 ArgPro-----GluAlaLeuGlyGlyThrLeuHisAsnThr-IleAlaGlyAs 49
DB 513 CGTCCCGCGGATACCGCGCGCGACGGTTCAAGGAGTTCGTCCACACACATCGCGCGTCC 572
QY 49 pLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuLeuArgValAl 69
DB 573 GCTCGTCCGCACTTCAGATCGCGCGCGCGGAGAACGAT-----TGGGGCGCGTACCGC 626
QY 69 aGluLeuAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaI 89
DB 627 CGAGATCGTCCGACACCGCCCGCTTCGACATCGATCTCGCGCACTCGATCCGACGCGT 686
QY 89 eGlnValGlySerProThrHisPro-----SerTyrProSer-- 101
DB 687 CATGGTAGACGATGATACGGGACCGGTAGCGGACCGGCGGCAACTACTCCCGATG 746
QY 102 -----GlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAl 116
DB 747 GCGGACGCGGACATGGAGCCCTCAACGGTCAACGCGACGCGACGCTCGTCTCGGAT 806
QY 116 aLeuLeuGlyLeuAspArgGlyGlyGluCysPheProAsnProValPheProSerAspAs 136
DB 807 CCTCTGACCATCATCGTGGCGGTGAG-----AGCAA 839
```



Score: 91.50 Matches: 50  
Percent Similarity: 39.66% Conservative: 21  
Best Local Similarity: 27.93% Mismatches: 74  
Query Match: 8.81% Indels: 34  
DB: 4 Gaps: 7

US-10-691-383-2\_COPY\_435\_632 (1-198) x AAF22877 (1-10480)

Qy 14 AlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnValHisArgPheAla 33  
Db 2951 GCCGGGGTTCGGCGCGCTACCAACCAACACCGGCGCCCTCATCGAGAAAT 3010  
Qy 34 ArgPro-----GluAlaLeuGlyThrLeuHisThr-IleAlaGlyAs 49  
Db 3011 CGTCCGGCGATCACCAGCGCGGTTCAAGGAGTTTCGTCACACACATCGCGGTC 3070  
Qy 49 pLeuAspAlaAspPheAspIleSerLeuGluAsnAspGluLeuLeuValAl 69  
Db 3071 GCTCGCTCGGCTTCAGATCGCGCGCGGAGAACAT-----TGGGGCGGTACCGC 3124  
Qy 69 aGluileAsnAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAla 89  
Db 3125 CGAGATCGTCGACACCCCTTCGACATCGATCTCGCGCACTCGATCCGACAGCGT 3184  
Qy 89 eGlnValGlySerProThrHisPro-----SerTyrProSer-- 101  
Db 3185 CATGGTGAAGACGATGACGCGGAGCGGTAGCGAGCGGCAATATCTCCGGATG 3244  
Qy 102 -----GlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuVal 116  
Db 3245 CGCGCAGCGGACATGGAGCCCTCAACGGTCAAGCGGCAACGCGCTCGCTCGGAT 3304  
Qy 116 aLeuileGlyLeuAspArgGlyGlyGluCysPheProAsnProValPheProSerAsp 136  
Db 3305 CCTCTCGACCATCAGCTGGCGGTGAG-----AGCAA 3337  
Qy 136 pGlyLeuGluLeuLeuAsnPheGluGlyAlaCysLeuThrTyrGluGlyLeuLeu 155  
Db 3338 CGGCTCCGCTCTCTCCACAGCAGCTCGATCGGCAAGATTTTCAAGAGCAGAACACGA 3397  
Qy 156 -----LysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyHisTyr 172  
Db 3398 CGTGCACCTCGCTCGGTGTGGCTTCGCGCG-----GGCATCGGCTAC 3443

RESULT 14  
ABL28263  
ID ABL28263 standard; DNA; 1191 BP.  
XX AC ABL28263;  
XX 26-MAR-2002 (first entry)  
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 36262.  
XX KW Drosophila; developmental biology; cell signalling; insecticide;  
XX OS Drosophila melanogaster.  
XX PN WO200171042-A2.  
XX PD 27-SEP-2001.  
XX PF 23-MAR-2001; 2001WO-US009231.  
XX PR 23-MAR-2000; 2000US-0191637P.  
XX PR 11-JUL-2000; 2000US-00614150.  
XX PA (PEKE ) PE CORP NY.  
XX PI Venter JC, Adams M, Li PWD, Myers EW;  
XX DR WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
PS Claim 1; SEQ ID NO 36262; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins (AB161737-  
CC AB172072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences)  
XX  
SQ Sequence 1191 BP; 281 A; 312 C; 326 G; 272 T; 0 U; 0 Other;

Alignment Scores: 2.01 Length: 1191  
Pred. No.: 88.50 Matches: 43  
Score: 38.89% Conservative: 20  
Percent Similarity: 26.54% Mismatches: 60  
Best Local Similarity: 8.52% Indels: 39  
Query Match: 4 Gaps: 6  
DB:

US-10-691-383-2\_COPY\_435\_632 (1-198) x ABL28263 (1-1191)

Qy 48 GlyAspLeuAspAlaAspPheAspIleSerLeuLeuGlnAsnAspGluLeuLeuLysArg 67  
Db 193 GGTAAATTTGGATGCCCTCTCTCGAGCATCTTTCGCAACGAACTTTTGGCAACCGG 252  
Qy 68 ValAlaGluLeuAsnAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuProGln 87  
Db 253 GTC---GAATTTATCCCAACAGCTGCTCAATCTGCAGCGCGCAGCTGCTCCCGCT 309  
Qy 88 AlaileGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGlnAsn 107  
Db 310 GGCCAGGGCGATGGATCCCATACATCCGAGTGGCCCAATCGGTGCTCCTTCGAC 369  
Qy 108 GlyAlaPheAlaThrValLeuLysAlaLeu----- 117  
Db 370 GGCGCGGAGCAAAATGTGATGAAGACCTTCTGCAGGTGCACAATTTGCCTGAGGCTC 429  
Qy 118 -----IleGlyLeuAspArgGlyGlyGluCysPheProAsnProValPheProSerAsp 135  
Db 430 GAGCGCTATGAGCCGACAACTCGGAGGAGAAATTACGACAA-----GAATCCAGC 480  
Qy 136 AspGlyLeuGluLeuLeuAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluLeuAsn 155  
Db 481 GATGGCATG-----CTGGGGGATATCTAC 504  
Qy 156 LysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyHisTyrArgPheAsp 175  
Db 505 GAGCAGCGCTGGAGATGGCCATTGA-----TGTATCTACAATTTGATGAT 552  
Qy 176 GlyileGlnLysLeuLeuGlyGluThrIleThrValargThrLeuHisGlnGluLeu 195  
Db 553 GGAATC-----ACGAAACCTTCGCACCATAGCCCGCTTCTTCG 591  
Qy 196 MetThr 197  
Db 592 GTGACC 597

RESULT 15  
AAF81756  
ID AAF81756 standard; cDNA; 2117 BP.  
XX AC AAF81756;  
XX 12-JUN-2001 (first entry)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 19, 2004, 04:03:44 ; Search time 237 Seconds  
(without alignments)  
3364.964 Million cell updates/sec

Title: US-10-691-383-2\_COPY\_435\_632  
Perfect score: 1039  
Sequence: 1 VNFSTSHYFLRIGAEALQK.....GLLGLGTTVTRLHQLMTF 198

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3327077 seqs, 2523723180 residues

Total number of hits satisfying chosen parameters: 6654154

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US10691383/runat\_17092004\_102657\_1795/app\_query.fasta\_1.1230  
-DB=Published Applications NA -Qfmt=fastap -SUFFIX=rnpb -MINWATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blowsum62  
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MTN=0 -ALIGN=15 -MODE=LOCAL -OUTPMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10691383 @CGN 1.1 776 @runat\_17092004\_102657\_1795  
-NCPU=6 -ICPU=3 -NO MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
13: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:  
17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:  
19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Score Match Length DB ID Description

RESULT 1									
US-09-840-762A-1									
; Sequence 1, Application US/09840762A									
; Patent No. US20020035245A1									
; GENERAL INFORMATION:									
; APPLICANT: Vreeland, Valerie									
; APPLICANT: Ng, Kwan L.									
; TITLE OF INVENTION: The Regents of the University of California									
; FILE REFERENCES: 023070-08710005									
; CURRENT APPLICATION NUMBER: US/09/840,762A									
; CURRENT FILING DATE: 2001-04-23									
; PRIOR APPLICATION NUMBER: 09/151,189									
; PRIOR FILING DATE: 2001-04-23									
; NUMBER OF SEQ ID NOS: 11									
; SOFTWARE: Patent in Ver. 2.0									
; SEQ ID NO 1									
; LENGTH: 2931									
; TYPE: DNA									
; ORGANISM: Fucus distichus									
; FEATURE:									
; NAME/KEY: CDS									
; LOCATION: (228)..(2258)									

ALIGNMENTS

1	1039	100.0	2931	9	US-09-840-762A-1	Sequence 1, Appli
2	1039	100.0	2931	17	US-10-691-383-1	Sequence 1, Appli
3	101.5	9.8	597	13	US-10-282-122A-41781	Sequence 41781, A
4	101	9.7	5828	9	US-09-430-029-1	Sequence 1, Appli
5	99	9.5	51	9	US-09-840-762A-3	Sequence 3, Appli
6	99	9.5	51	17	US-10-691-383-3	Sequence 3, Appli
7	93	9.0	1575	15	US-10-156-761-7467	Sequence 7467, Ap
8	93	9.0	9025608	15	US-10-156-761-1	Sequence 1, Appli
9	91.5	8.8	1161	15	US-10-273-051-3	Sequence 3, Appli
10	91.5	8.8	1161	16	US-10-129-518-3	Sequence 3, Appli
11	91.5	8.8	10480	15	US-10-273-051-13	Sequence 13, Appli
12	91.5	8.8	10480	16	US-10-129-518-13	Sequence 13, Appli
13	89	8.6	1282	13	US-10-425-114-22562	Sequence 22562, A
14	89	8.6	1506	17	US-10-767-701-12438	Sequence 12438, A
15	88.5	8.5	1944	13	US-10-169-395-14	Sequence 14, Appli
16	88.5	8.5	2117	9	US-09-965-529-53	Sequence 53, Appli
17	88.5	8.5	2117	10	US-09-969-680A-53	Sequence 53, Appli
18	88.5	8.5	2122	13	US-10-169-395-24	Sequence 24, Appli
19	88.5	8.5	2192	16	US-10-353-690-75	Sequence 75, Appli
20	88.5	8.5	2192	16	US-10-295-027-1120	Sequence 1120, Ap
21	88	8.5	819	13	US-10-282-122A-30580	Sequence 30580, A
22	88	8.5	1894	16	US-10-369-493-36021	Sequence 36021, A
23	87.5	8.4	1815	9	US-09-832-616-1	Sequence 1, Appli
24	87.5	8.4	1815	9	US-09-832-616-1	Sequence 1, Appli
25	87.5	8.4	2706	16	US-10-320-797-2320	Sequence 2320, Ap
26	86	8.3	579	13	US-10-282-122A-27464	Sequence 27464, A
27	86	8.3	1485	13	US-10-425-114-26985	Sequence 26985, A
28	85.5	8.2	3771	9	US-09-854-356-11	Sequence 11, Appli
29	85.5	8.2	12279	13	US-10-363-616-80	Sequence 80, Appli
30	84.5	8.1	6197	15	US-10-085-959-114	Sequence 114, App
31	84	8.1	504	9	US-09-738-626-3176	Sequence 3176, Ap
32	84	8.1	1359	15	US-10-156-761-738	Sequence 738, App
33	84	8.1	3309400	9	US-09-738-626-1	Sequence 1, Appli
34	83.5	8.0	525	13	US-10-282-122A-16647	Sequence 16647, A
35	83.5	8.0	1161	15	US-10-273-051-30	Sequence 30, Appli
36	83.5	8.0	2310	17	US-10-129-518-30	Sequence 30, Appli
37	83.5	8.0	2310	17	US-10-115-635-153	Sequence 153, App
38	83.5	8.0	2317	15	US-10-097-340-94	Sequence 94, Appli
39	83.5	8.0	2317	16	US-10-093-463-119	Sequence 119, App
40	82.5	7.9	659	13	US-10-425-114-35812	Sequence 35812, A
41	82.5	7.9	1839	13	US-10-424-599-43078	Sequence 43078, A
42	82	7.9	1026	16	US-10-369-493-24328	Sequence 24328, A
43	82	7.9	18529	15	US-10-198-846-12599	Sequence 12599, A
44	82	7.9	92638	17	US-10-450-826-3	Sequence 3, Appli
45	81.5	7.8	1797	9	US-09-833-102-1	Sequence 1, Appli

```
; OTHER INFORMATION: vanadium bromoperoxidase
US-09-840-762A-1

Alignment Scores:
Pred. No.: 1,06e-134 Length: 2931
Score: 1039.00 Matches: 198
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-10-691-383-2_COPY_435_632 (1-198) x US-09-840-762A-1 (1-2931)
Qy 1 ValAsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArg 20
Db 1530 GTCAACTTCGGCAGCTCTCACTACTTCAGATTGATAGTGCCCGGAGCTGGCGACGCT 1589
Qy 21 AlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGly 40
Db 1590 GCCTCGTGTACCAAAAGTGGCAGGTGCATCGATTTCAGATGCAGACTTCGACATCTCCCTTTTGAA 1709
Qy 41 ThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGlu 60
Db 1650 ACCCTCCACAACACCACCGGGGATCTAGATGCAGATTCGACATCTCCCTTTTGAA 1709
Qy 61 AsnAspGluLeuLeuLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGlu 80
Db 1710 AATGATGAGCTCTTGAACGTGTGGCGGAGATAAATGCGGCGAGATCCCAACACGAG 1769
Qy 81 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 100
Db 1770 GTCACTACCTTTCTTCACAAGCTATCCAGTGGGATCGCAACGACACCTTCTTACCCG 1829
Qy 101 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeu 120
Db 1830 TCCGGCCACGCTACCCAAAATGGAGCATTTGCCACAGTTCTGAAGGCCCTCATTTGGCCTA 1889
Qy 121 AspArgGlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeu 140
Db 1890 GATCGGGGAGTGAGTGCTTCCCTAACCCCGTGTCCCAAGCGATGACGGCTCGGCTCAAC 2009
Qy 141 IleAsnPheGluGlyAlaCysLeuThrTyrGluGlyIleHisTyrArgPheAspGlyIleGlnGlyLeu 180
Db 2010 GTGCGATTGGGAGCAGATGCTGGGCATCCACTATCGGTTCGACGGTATCCAAAGGCCTA 2069
Qy 181 LeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThrPhe 198
Db 2070 CTCTCGGAGAGACATCACTGTACGAACACTTCACCGAGGAGCTGATGAGCTTC 2123

RESULT 2
US-10-691-383-1
; Sequence 1, Application US/10691383
; Publication No. US20040110260A1
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; TITLE OF INVENTION: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases
; FILE OF INVENTION: and Their Uses
; FILE REFERENCE: 023070-0871110US
; CURRENT APPLICATION NUMBER: US/10/691,383
; CURRENT FILING DATE: 2003-10-21
; PRIOR APPLICATION NUMBER: US/09/596,794
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: US 09/151,189
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2931

; TYPE: DNA
; ORGANISM: Fucus distichus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (228)..(2258)
; OTHER INFORMATION: vanadium bromoperoxidase
US-10-691-383-1

Alignment Scores:
Pred. No.: 1,06e-134 Length: 2931
Score: 1039.00 Matches: 198
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 17 Gaps: 0

US-10-691-383-2_COPY_435_632 (1-198) x US-10-691-383-1 (1-2931)
Qy 1 ValAsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArg 20
Db 1530 GTCAACTTCGGCAGCTCTCACTACTTCAGATTGATAGTGCCCGGAGCTGGCGACGCT 1589
Qy 21 AlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGly 40
Db 1590 GCCTCGTGTACCAAAAGTGGCAGGTGCATCGATTTCAGATGCAGACTTCGACATCTCCCTTTTGAA 1709
Qy 41 ThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGlu 60
Db 1650 ACCCTCCACAACACCACCGGGGATCTAGATGCAGATTCGACATCTCCCTTTTGAA 1709
Qy 61 AsnAspGluLeuLeuLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGlu 80
Db 1710 AATGATGAGCTCTTGAACGTGTGGCGGAGATAAATGCGGCGAGATCCCAACACGAG 1769
Qy 81 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 100
Db 1770 GTCACTACCTTTCTTCACAAGCTATCCAGTGGGATCGCAACGACACCTTCTTACCCG 1829
Qy 101 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeu 120
Db 1830 TCCGGCCACGCTACCCAAAATGGAGCATTTGCCACAGTTCTGAAGGCCCTCATTTGGCCTA 1889
Qy 121 AspArgGlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeu 140
Db 1890 GATCGGGGAGTGAGTGCTTCCCTAACCCCGTGTCCCAAGCGATGACGGCTCGGCTCAAC 2009
Qy 141 IleAsnPheGluGlyAlaCysLeuThrTyrGluGlyIleHisTyrArgPheAspGlyIleGlnGlyLeu 180
Db 2010 GTGCGATTGGGAGCAGATGCTGGGCATCCACTATCGGTTCGACGGTATCCAAAGGCCTA 2069
Qy 181 LeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThrPhe 198
Db 2070 CTCTCGGAGAGACATCACTGTACGAACACTTCACCGAGGAGCTGATGAGCTTC 2123

RESULT 3
US-10-282-122A-41781
; Sequence 41781, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
```





```
QY 84 LeuLeuProGlnAlaIleGlnValGlySerProThrHis---ProSerTyrProSerGly 102
|||:|||||
Db 268 CTGTCACCGCAGATCCGCCACCTCACCCGCGAGCGGCACACACCTCTCTTCCCGTCGGGA 327
|||:|||||

QY 103 HisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArg 122
|||:|||||
Db 328 CACTCGCCCTCGCGCGCGGCTTTCGACCGGC-----GTCGCCCTCGAGTCC 375
|||:|||||

QY 123 GlyGlyGluCysPheProAsnProValPheProSerAspGlyLeuGluLeuLeuAsn 142
|||
Db 376 ACCGGC-----381

QY 143 PheGluGlyAlaCysLeuThrTyrGluGlyLeuLeuAsnLysLeuAlaValAla 162
|||:|||||
Db 382 -----TACGGCGCAGTCTGCGCGCGCTGCGCGCGGCGGCGTCCGCC 420
|||:|||||

QY 163 PheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeu 182
|||:|||||
Db 421 TTCTCTCGGCTCTATGTCGGCGTGCACCTACCCCGGGACGTCCTGCGCGGTATGGCGATC 480
|||:|||||

QY 183 Gly 183
|||
Db 481 GGC 483

RESULT 8
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.: 1.35e+04 Length: 9025608
Score: 93.00 Matches: 28
Percent Similarity: 39.60% Conservative: 12
Best Local Similarity: 27.72% Mismatches: 31
Query Match: 8.95% Indels: 30
DB: 15 Gaps: 3

US-10-691-383-2_COPY_435_632 (1-198) x US-10-156-761-1 (1-9025608)

QY 84 LeuLeuProGlnAlaIleGlnValGlySerProThrHis---ProSerTyrProSerGly 102
|||:|||||
Db 8921985 CTGTCACCGCAGATCCGCCACCTCACCCGCGAGCGGCACACACCTCTCTTCCCGTCGGGA 8922044
|||:|||||

QY 103 HisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArg 122
|||:|||||
Db 8922045 CACTCGCCCTCGCGCGCGGCTTTCGACCGGC-----GTCGCCCTCGAGTCC 8922092
|||:|||||
```

```
QY 123 GlyGlyGluCysPheProAsnProValPheProSerAspGlyLeuGluLeuLeuAsn 142
|||
Db 8922093 ACCGGC-----8922098

QY 143 PheGluGlyAlaCysLeuThrTyrGluGlyLeuLeuAsnLysLeuAlaValAla 162
|||:|||||
Db 8922099 -----TACGGCGCAGTCTGTCGGCGTGCACCTACCCCGGGACGTCCTGCGCGGTATGGCGATC 8922137
|||:|||||

QY 163 PheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeu 182
|||:|||||
Db 8922138 TTCTCTCGGCTCTATGTCGGCGTGCACCTACCCCGGGACGTCCTGCGCGGTATGGCGATC 8922197
|||:|||||

QY 183 Gly 183
|||
Db 8922198 GGC 8922200

RESULT 9
US-10-273-051-3
; Sequence 3, Application US/10273051
; Publication No. US20030157673A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Mario W.
; APPLICANT: Chen, Qiong
; APPLICANT: Gibson, Katherine J.
; APPLICANT: Kostichka, Kristy N.
; APPLICANT: Thomas, Stuart M.
; APPLICANT: Nagatajan, Vasantha
; TITLE OF INVENTION: Genes Involved in Cyclododecanone Degradation Pathway
; FILE REFERENCE: BC1023 US NA
; CURRENT APPLICATION NUMBER: US/10/273,051
; CURRENT FILING DATE: 2002-10-17
; PRIOR APPLICATION NUMBER: 60/170,214
; PRIOR FILING DATE: 1999-12-10
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 3
; LENGTH: 1161
; TYPE: DNA
; ORGANISM: Rhodococcus ruber
; FEATURE:
US-10-273-051-3

Alignment Scores:
Pred. No.: 0.0297 Length: 1161
Score: 91.50 Matches: 50
Percent Similarity: 39.66% Conservative: 21
Best Local Similarity: 27.93% Mismatches: 74
Query Match: 8.81% Indels: 34
DB: 15 Gaps: 7

US-10-691-383-2_COPY_435_632 (1-198) x US-10-273-051-3 (1-1161)

QY 14 AlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAla 33
|||:|||||
Db 453 GCCCGCGGGTTTCGGCGCGGCTTACCACGCCCAACACAGGCCGCCCTCTATCCAGAGAAAT 512
|||:|||||

QY 34 ArgPro-----GluAlaLeuGlyGlyThrLeuHisAsnThr-IleAlaGlyAs 49
|||:|||||
Db 513 CGTCCCGCGCATCACCGCGCGCGGCTTCAAGGAGTTGTCACACACACATCCCGGTC 572
|||:|||||

QY 49 pLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLysArgValAl 69
|||:|||||
Db 573 GCTCGCTCGCGACTTCCAGATCGCGCGCGCGAGACGAT-----TGGGCGCGTACCGC 626
|||:|||||

QY 69 aGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaI 89
|||:|||||
Db 627 CGAGATCGTGCACACACCCCGCTTCGACATCGATCTCGCGCGCATCGATCCGACACGCT 686
|||:|||||

QY 89 eGlnValGlySerProThrHisPro-----SerTyrProSer-- 101
|||:|||||
Db 687 CATGGTGAACACGATGACGGGACCGGTAGCCCGACCGGACCGCGGCCNACTACTCCGGATG 746
|||:|||||
```





Qy 136 pGlyLeuGluLeuLeuLeuPheGluGlyAlaCysLeuThrTyrGluGlyGluLeuLeuAen-- 155  
Db 3338 CGGCTCGGCTCTCTCCACAGCACCTCGATCGGCAAGATTTTCGAGAGCAGACAACGA 3397  
Qy 156 ----LysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyLeuHisTyr 172  
Db 3398 CGTCGACCTCGTCTCGGTGTCGCTTCGCGCGC-----GGCATCGGCTAC 3443

## RESULT 12

US-10-129-518-13  
; Sequence 13, Application US/10129518  
; Publication No. US20030215930A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. du Pont de Nemours and Company  
; TITLE OF INVENTION: Genes Involved in Cyclododecanone Degradation Pathway  
; FILE REFERENCE: BC1023 PCT  
; CURRENT APPLICATION NUMBER: US/10/129,518  
; CURRENT FILING DATE: 2002-11-04  
; PRIOR APPLICATION NUMBER: 60/170,214  
; PRIOR FILING DATE: 1999-12-10  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 13  
; LENGTH: 10480  
; TYPE: DNA  
; ORGANISM: Rhodococcus ruber  
US-10-129-518-13

Alignment Scores:  
Pred. No.: 0.821 Length: 10480  
Score: 91.50 Matches: 50  
Percent Similarity: 39.66% Conservative: 21  
Best Local Similarity: 27.93% Mismatches: 74  
Query Match: 8.81% Indels: 34  
DB: 16 Gaps: 7

US-10-691-383-2\_COPY\_435\_632 (1-198) x US-10-129-518-13 (1-10480)

Qy 14 AlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAla 33  
Db 2951 GCCGGGGTTCGGCCGCGGTACACCGCCACCAACACCGGGCCACCTCATCGGAGAAAT 3010  
Qy 34 ArgPro-----GluAlaLeuGlyGlyThrLeuHisAenThr-IleAlaGlyAs 49  
Db 3011 CGTCGCGCGCATCACCGCCGCGACCGTTCAAGAGAGTTCGTCCACACACACATCGCGGTC 3070  
Qy 49 pLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLysArgValAl 69  
Db 3071 GCTCGTGGCGGACTTCAGATCGCGCGCGCGAGAACGAT-----TGGGGCGGTACCGC 3124  
Qy 69 aGluLeuAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaI 89  
Db 3125 CGAGATCGTCGCACCCCGCTTCGACATCGATCTCGCGCATCGATCCGACACGCT 3184  
Qy 89 eGlnValGlySerProThrHisPro-----SerTyrProSer-- 101  
Db 3185 CATGGTGAAGACGATGACGGACCGGATGCGGACCGGACCGGACCGGACCGGATG 3244  
Qy 102 -----GlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAl 116  
Db 3245 CGCGCACCGCACATGAGGAGCCCTCAACGGTCAAGCGCAACGCGCTCGCTCGGAT 3304  
Qy 116 aLeuLeuGlyLeuAspArgGlyGlyGlyCysPheProAsnProValPheProSerAspAs 136  
Db 3305 CCTTCGACCATCAGCTGGGGGTGAG-----AGCAA 3337  
Qy 136 pGlyLeuGluLeuLeuLeuPheGluGlyAlaCysLeuThrTyrGluGlyGluLeuAen-- 155  
Db 3338 CGGCTCGGCTCTCTCCACAGCACCTCGATCGGCAAGATTTTCGAGAGCAGACAACGA 3397  
Qy 156 ----LysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyLeuHisTyr 172  
Db 3398 CGTCGACCTCGTCTCGGTGTCGCTTCGCGCGC-----GGCATCGGCTAC 3443

## RESULT 13

US-10-425-114-22562  
; Sequence 22562, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(53313)B  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 22562  
; LENGTH: 1282  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3587-278-C11\_FLI  
US-10-425-114-22562

Alignment Scores:  
Pred. No.: 0.0775 Length: 1282  
Score: 89.00 Matches: 53  
Percent Similarity: 34.86% Conservative: 23  
Best Local Similarity: 24.31% Mismatches: 64  
Query Match: 8.57% Indels: 78  
DB: 13 Gaps: 11

US-10-691-383-2\_COPY\_435\_632 (1-198) x US-10-425-114-22562 (1-1282)

Qy 11 LeuLeuGlyAlaAlaGluLeuAlaGlnArg---AlaSerCysTyrGlnLysTrpGlnVal 29  
Db 166 CTCACCGGATCGCTTCGCGCGGAGCGCGGGCGTGTCTCTCGACCCATCCGGGTG 225  
Qy 30 HisArgPheAla---ArgProGluAlaLeuGlyGlyThrLeuHisAenThrIleAlaGly 48  
Db 226 GTTCAACTCTCTTGGCGCCCGCGGGCGTTC-----TTGCACAAAGGTTTCTCTGTG 276  
Qy 49 AspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLysArgVal 68  
Db 277 GATCGGAGTGCACACCTCATCGCGCTGCCACGACGAGTGGAGAGTCTATGTG 336  
Qy 69 AlaGluLeuAsnAlaAlaGlnAsnProAsnAsnGluVal----- 81  
Db 337 GCGGACAAACAGTCCGTTAAGAGTGTCCAGAGTGGCGGACAGTTCGGCATGTT 396  
Qy 82 -----Thr 82  
Db 397 CTCGAGAAGAGCAGGATGAAGTAGTAACGAGGATAGAGAGGATATCTGTTGACA 456  
Qy 83 TyrLeuLeuPro-----GlnAlaIleGlnVal-----GlySerPro 94  
Db 457 TTTCTCCACAGAGATGTTGAGGACCATTCAGATATTACACTACAGAGTGTGAGAAA 516  
Qy 95 ThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAla----- 109  
Db 517 TATGAACCCCACTACGACTACTTCCATGACAAAAATAATCAGGCCCTCGGGCGGCATCG 576  
Qy 110 PheAlaThrValLeuLysAlaLeuLeuGlyLeuAspArgGlyGlyGlu---CysPhePro 128  
Db 577 ATTGCCCATGTGTGTATGTATCTATCAAAATGTGAGAAAGGTGGAGAGACATCTTCCCC 636  
Qy 129 AsnPro-----ValPheProSerAspAsp----- 136  
Db 637 ATGCGAGGGGGAAGCTATTGCAACCCACGACGACACACTTGGTGTGTTGTCGAGAAC 696  
Qy 137 -----GlyLeuGluLeuIleAsnPhe----- 143

Db 697 GGATATGCACTTAACCGGTAAAGGTTGATGCCCTGCTGTTCTTCAGTCTCCACCCCTGAT 756  
Qy 144 -----GluGlyAlaCysLeuThrTyrGluGlyGlu 153  
Db 757 TCAACAACAGACTCTGACAGCTTGACAGCTAGTGCCTCCGTCATCGAAGGCCAG 810  
RESULT 14  
US-10-767-701-12438  
; Sequence 12438, Application US/10767701  
; Publication No. US20040172684A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement  
; FILE REFERENCE: 38-21(53535)B  
; CURRENT APPLICATION NUMBER: US/10/767,701  
; CURRENT FILING DATE: 2004-01-29  
; NUMBER OF SEQ ID NOS: 63128  
; SEQ ID NO 12438  
; LENGTH: 1506  
; TYPE: DNA  
; ORGANISM: Sorghum bicolor  
; FEATURE:  
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS9661\_1  
US-10-767-701-12438  
Alignment Scores:  
Pred. No.: 0.0988 Length: 1506  
Score: 89.00 Matches: 54  
Percent Similarity: 35.62% Conservative: 24  
Best Local Similarity: 24.66% Mismatches: 61  
Query Match: 8.57% Indels: 80  
DB: 17 Gaps: 13  
US-10-691-383-2\_COPY\_435\_632 (1-198) x US-10-767-701-12438 (1-1506)  
Qy 11 LeuileGlyAlaAlaGluLeuAlaGlnArg---AlaSerCysTyrGlnLysTrpGlnVal 29  
Db 433 CTTCCGGCTCGCGCTCTGGCGGAGCGGGGGCTGTTCTTCGACCCCTCCCGCGTG 492  
Qy 30 HisArgPheAla---ArgProGluAlaLeuGlyThrLeuHisThrIleAlaGly 48  
Db 493 GTCAACTCTCGGGCCGCCAGGCGCTTC-----CTGCACAAAGGTTTCCTGTCG 543  
Qy 49 AspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLys---Arg 67  
Db 544 GATCGGAGTGGAC---CACCTCATGCTGTCGCCAAGACAACTGGAGAGTCTATG 600  
Qy 68 ValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluVal----- 81  
Db 601 GTAGCGGACAAACGAGTCCGGTAAGAGCGTCCAGAGCGAGGTGGCGCACCACTCCGGCATG 660  
Qy 81 ----- 81  
Db 661 TTCTGGAGAGAGCAGGATGAAGTAGTAGAGGGAATAGAGGAGGATAGTGTGG 720  
Qy 82 ThrTyrLeuLeuPro-----GlnAlaIleGlnVal-----GlySer 93  
Db 721 ACATTCTTCCACAGAGAATGGTGAATCCATTACAGATATTACACTACAGAAATGGTGAG 780  
Qy 94 ProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAla----- 109  
Db 781 AAATACGAACCCCACTATGACTACTTCCATGACAAAAATAATCAAGCCCTGGGTGGCCAT 840  
Qy 110 ---PheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGlu---CysPhe 127  
Db 841 CGCATGCCACTGTGCTCATGTACTTATCTAATGTTGAGAGGTTGAGACCATCTTC 900  
Qy 128 ProAsnPro-----ValPheProSerAspAsp----- 136  
::: ||| |||||

Db 901 CCAATGCAGAGGGAAGCTATTACAACCAAGGACGACACATGGTCTGATTGTGCAAGA 960  
Qy 137 -----GlyLeuGluLeuIleAsnPhe----- 143  
Db 961 AATGGATATGCAGTTAAACCGGTAAAGGGTGATGCGCTGCTGTTCTTCAGTCTCCACCT 1020  
Qy 144 -----GluGlyAlaCysLeuThrTyrGluGlyGlu 153  
Db 1021 GATGCAACAACAGACTCTGAGAGCTTGACAGCTAGTGCCTCCGTCATCGAAGGCCAG 1077  
RESULT 15  
US-10-169-395-14  
; Sequence 14, Application US/10169395  
; Publication No. US20040034192A1  
; GENERAL INFORMATION:  
; APPLICANT: KATO, Seishi  
; APPLICANT: KIMURA, Tomoko  
; TITLE OF INVENTION: HUMAN PROTEINS HAVING HYDROPHOBIC DOMAINS AND DNAS ENCODING  
; TITLE OF INVENTION: THESE PROTEINS  
; FILE REFERENCE: 01997.015100.US  
; CURRENT APPLICATION NUMBER: US/10/169,395  
; CURRENT FILING DATE: 2002-11-29  
; PRIOR APPLICATION NUMBER: JP 2000-585  
; PRIOR FILING DATE: 2000-01-06  
; PRIOR APPLICATION NUMBER: JP 2000-588  
; PRIOR FILING DATE: 2000-01-06  
; PRIOR APPLICATION NUMBER: JP 2000-2299  
; PRIOR FILING DATE: 2000-01-11  
; PRIOR APPLICATION NUMBER: JP 2000-26862  
; PRIOR FILING DATE: 2000-02-03  
; PRIOR APPLICATION NUMBER: JP 2000-58367  
; PRIOR FILING DATE: 2000-03-03  
; PRIOR APPLICATION NUMBER: PCT/JP00/09359  
; PRIOR FILING DATE: 2000-12-28  
; NUMBER OF SEQ ID NOS: 150  
; SEQ ID NO 14  
; LENGTH: 1944  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-169-395-14  
Alignment Scores:  
Pred. No.: 0.171 Length: 1944  
Score: 88.50 Matches: 38  
Percent Similarity: 41.79% Conservative: 18  
Best Local Similarity: 28.36% Mismatches: 39  
Query Match: 8.52% Indels: 39  
DB: 13 Gaps: 7  
US-10-691-383-2\_COPY\_435\_632 (1-198) x US-10-169-395-14 (1-1944)  
Qy 15 AlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArg 34  
Db 280 GCCCGCTCAGTCCGCCCGCGCTCTGTATC-----CTCAGCAAC 318  
Qy 35 ProGluAlaLeuGlyThrLeuHisAsnThrIleAlaGly-----AspLeuAspAla 52  
Db 319 CCCAG-----GGCACCTGTGAGGACACTCGGGCTGGCCCTCTCATGCA 369  
Qy 53 AspPheAspIleSerLeuLeuGluAsnAspGlu-----Leu 64  
Db 370 GACCACCTCTCGCCCTGCTCGAGAGCCCAAGGCCCTGACCCCGGCGCTGAGCTGGCTG 429  
Qy 65 LeuLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsn-----GluVal 81  
Db 430 CTGCAGAGGATGAGCCCGGCTGCGGCGCAGACCCCAAGACGCGCTCGTAGATATC 489  
Qy 82 ThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSer 101  
Db 490 CCTCAGCTCTGAGAGAGCGGTGGGGGGGGGGCTCG----- 528  
Qy 102 GlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAsp 121  
::: ||| |||||

Db 529 -----GGCAGTGTGTGGCGGCTCCTGGCTGCCCTGTGACCATGTC 570  
Qy 122 ArgGlyGlyGluCysPhe-----ProAsnProValPhe 132  
Db 571 AGGAGCGGGTCTTGCTTCCACGCCCTTGCCCGAGCCCTCAGTAC 612

Search completed: September 19, 2004, 11:18:46  
Job time : 2394 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 19, 2004, 04:00:38 ; Search time 42.5904 Seconds  
(without alignments)  
2579.931 Million cell updates/sec

Title: US-10-691-383-2\_COPY\_435\_632  
Perfect score: 1039  
Sequence: 1 VNFGTSHYFLRIGABLAQR.....GLLGLSTITVTLHQLMTF 198

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO spool/US10691383/runat 17092004 102657 1766/app query.fasta\_1.1230  
-DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=Blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10691383 @CGN 1 93 @runat 17092004 102657 1766 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1039	100.0	2931	3	US-09-151-189-1
2	1039	100.0	2931	4	US-09-596-794-1
3	108	10.4	4403765	3	US-09-103-840A-2
4	108	10.4	411529	3	US-09-103-840A-1
5	99	9.5	51	3	US-09-151-189-3
6	99	9.5	51	4	US-09-596-794-3
7	91.5	8.8	1161	4	US-09-732-615-3
8	91.5	8.8	10480	4	US-09-732-615-13
9	87.5	8.4	1815	4	US-09-832-496-1
10	87.5	8.4	1815	4	US-09-832-616-1
11	87	8.4	1812	4	US-09-832-615-1
12	87	8.4	1812	4	US-09-832-617-1

13	85.5	8.2	1848	4	US-09-252-991A-11564	Sequence 11564, A	
14	84.5	8.1	1358	4	US-09-180-109A-25	Sequence 25, Appl	
15	83.5	8.0	1161	4	US-09-732-615-30	Sequence 30, Appl	
16	83.5	8.0	2822	2	US-08-679-405-1	Sequence 1, Appli	
17	83.5	8.0	2822	2	US-08-842-799-1	Sequence 1, Appli	
18	83.5	8.0	2822	5	PCT-US96-11458-1	Sequence 1, Appli	
19	82.5	7.9	873	4	US-09-134-001C-436	Sequence 436, Appl	
20	82.5	7.9	936	4	US-09-489-039A-5842	Sequence 5842, Appl	
21	82.5	7.9	1664976	4	US-08-916-421B-1	Sequence 1, Appli	
22	81.5	7.8	1797	4	US-09-832-441-1	Sequence 1, Appli	
23	81.5	7.8	1797	4	US-09-833-102-1	Sequence 1, Appli	
24	80.5	7.7	1344	4	US-09-252-991A-16526	Sequence 16526, A	
25	80.5	7.7	2631	4	US-09-252-991A-16070	Sequence 16070, A	
C	25	79	7.6	26270	4	US-09-717-364A-1	Sequence 1, Appli
26	79	7.6	4403765	3	US-09-103-840A-2	Sequence 2, Appli	
27	79	7.6	4411529	3	US-09-103-840A-1	Sequence 1, Appli	
28	79	7.6	726	4	US-09-252-991A-15963	Sequence 15963, A	
29	78.5	7.6	972	4	US-09-252-991A-16542	Sequence 16542, A	
C	30	78.5	7.6	1443	4	US-09-252-991A-16057	Sequence 16057, A
31	78.5	7.6	2094	4	US-09-489-039A-314	Sequence 314, Appl	
32	78.5	7.6	735	3	US-08-750-145A-23	Sequence 23, Appl	
33	76.5	7.4	735	3	US-08-975-698A-27	Sequence 27, Appl	
34	76.5	7.4	735	3	US-09-417-090-27	Sequence 27, Appl	
35	76.5	7.4	735	3	US-09-727-578-27	Sequence 27, Appl	
36	76.5	7.4	735	4	US-09-252-991A-11048	Sequence 11048, A	
37	75.5	7.3	1509	4	US-09-252-991A-11206	Sequence 11206, A	
C	38	75.5	7.3	1590	4	US-09-832-498-1	Sequence 1, Appli
39	75.5	7.3	1842	4	US-09-832-614A-1	Sequence 1, Appli	
40	75.5	7.3	1842	4	US-08-311-731A-137	Sequence 137, Appl	
41	75.5	7.3	40123	4	US-08-311-731A-137	Sequence 137, Appl	
42	75	7.2	924	3	US-08-952-089A-14	Sequence 14, Appl	
C	43	75	7.2	924	3	US-08-952-089A-29	Sequence 29, Appl
44	74.5	7.2	4146	3	US-08-952-089A-29	Sequence 29, Appl	
45	74.5	7.2	1863	4	US-09-252-991A-3225	Sequence 3225, Appl	
C	45	74.5	7.2	2154	4	US-09-252-991A-3346	Sequence 3346, Appl

ALIGNMENTS

RESULT 1

US-09-151-189-1  
; Sequence 1, Application US/09151189  
; Patent No. 6232457  
; GENERAL INFORMATION:  
; APPLICANT: Vreeland, Valerie  
; APPLICANT: Ng, Kwan L.  
; APPLICANT: The Regents of the University of California  
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses  
; FILE REFERENCE: 023070-087100US  
; CURRENT APPLICATION NUMBER: US/09/151,189  
; CURRENT FILING DATE: 1998-09-10  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 2931  
; TYPE: DNA  
; ORGANISM: Fucus distichus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (228)..(2258)  
; OTHER INFORMATION: vanadium bromoperoxidase  
US-09-151-189-1

Alignment Scores:  
Pred. No.: 4,94e-126 Length: 2931  
Score: 1039.00 Matches: 198  
Percent Similarity: 100.00% Conservativeness: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 3 Gaps: 0  
US-10-691-383-2\_COPY\_435\_632 (1-198) x US-09-151-189-1 (1-2931)

Qy 1 ValAspPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArg 20  
|||||

Db 1530 GTCAACTTCGGCAGCTCTCACTACTTCAGATTGATAGTGCCTGGCAGTGGCGCAGCGT 1589  
Qy 21 AlasSerCysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGly 40  
Db 1590 GCCTCGTGTATACCAAAAGTGGCAGGTGCATCGATTTCACAGCCCGCCGAGGCTCTCGGGGT 1649  
Qy 41 ThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGlu 60  
Db 1650 ACCCTCCACAAACACCATCGCGGGGATCTAGATGCAGATTCGACATCTCCCTCTTTGAA 1709  
Qy 61 AsnAspGluLeuLeuLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGlu 80  
Db 1710 AATGATGAGCTCTTGAAACGTGTGGCGGAGATAAATCGCGGCGAGATCCCAACACGAG 1769  
Qy 81 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 100  
Db 1770 GTCACTTACCTTCTTCACAAAGTATCCAAAGTGGATCGCAGACCCCTTCTTACCCG 1829  
Qy 101 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeu 120  
Db 1830 TCCGGCCACCTACTCCCAAAATGGAGCATTTGCCACAGTTCTGAAGGCCCTCATTTGGCCTA 1889  
Qy 121 AspArgGlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeu 140  
Db 1890 GATCGGGAGGTGAGTCTTCCCTAACCCCGTGTCCCAAGCGATGACGCGCTGGAATA 1949  
Qy 141 IleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsn 160  
Db 1950 ATCAACTTCGAAGGGCATGCTTACATATGAGGAGAGATCAACAAGCTCGCGGTCAAC 2009  
Qy 161 ValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeu 180  
Db 2010 GTCGCATTTGGAGGCGAGATGCTGGGCATCCACTATCGGTTCCAGCGTATCCAAAGGCTA 2069  
Qy 181 LeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThrPhe 198  
Db 2070 CTCTCGGAGAGACAATCACTGTACGAACATTCACCGAGGCTGATGAGCTTC 2123

## RESULT 2

US-09-596-794-1  
; Sequence 1, Application US/09596794  
; Patent No. 6656715  
; GENERAL INFORMATION:  
; APPLICANT: Vreeland, Valerie  
; APPLICANT: The Regents of the University of California  
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases  
; TITLE OF INVENTION: and Their Uses  
; FILE REFERENCE: 023070-0871110US  
; CURRENT APPLICATION NUMBER: US/09/596,794  
; CURRENT FILING DATE: 2000-06-19  
; PRIOR APPLICATION NUMBER: US 09/151,189  
; PRIOR FILING DATE: 1998-09-10  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 2931  
; TYPE: DNA  
; ORGANISM: Fucus distichus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (228)..(2258)  
; OTHER INFORMATION: vanadium bromoperoxidase  
US-09-596-794-1

Alignment Scores:  
Pred. No.: 4.94e-126 Length: 2931  
Score: 1039.00 Matches: 198  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-10-691-383-2\_COPY\_435\_632 (1-198) x US-09-596-794-1 (1-2931)

Qy 1 ValAsnPheGlyThrSerHisTyrPheArgLeuIleGlyValaAlaGluLeuAlaGlnArg 20  
Db 1530 GTCAACTTCGGCAGCTCTCACTACTTCAGATTGATAGTGCCTGGCAGTGGCGCAGCGT 1589  
Qy 21 AlasSerCysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGly 40  
Db 1590 GCCTCGTGTATACCAAAAGTGGCAGGTGCATCGATTTCACAGCCCGCCGAGGCTCTCGGGGT 1649  
Qy 41 ThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGlu 60  
Db 1650 ACCCTCCACAAACACCATCGCGGGGATCTAGATGCAGATTCGACATCTCCCTCTTTGAA 1709  
Qy 61 AsnAspGluLeuLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGlu 80  
Db 1710 AATGATGAGCTCTTGAAACGTGTGGCGGAGATAAATCGCGGCGAGATCCCAACACGAG 1769  
Qy 81 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 100  
Db 1770 GTCACTTACCTTCTTCACAAAGTATCCAAAGTGGATCGCAGACCCCTTCTTACCCG 1829  
Qy 101 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeu 120  
Db 1830 TCCGGCCACCTACTCCCAAAATGGAGCATTTGCCACAGTTCTGAAGGCCCTCATTTGGCCTA 1889  
Qy 121 AspArgGlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeu 140  
Db 1890 GATCGGGAGGTGAGTCTTCCCTAACCCCGTGTCCCAAGCGATGACGCGCTGGAATA 1949  
Qy 141 IleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsn 160  
Db 1950 ATCAACTTCGAAGGGCATGCTTACATATGAGGAGAGATCAACAAGCTCGCGGTCAAC 2009  
Qy 161 ValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeu 180  
Db 2010 GTCGCATTTGGAGGCGAGATGCTGGGCATCCACTATCGGTTCCAGCGTATCCAAAGGCTA 2069  
Qy 181 LeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThrPhe 198  
Db 2070 CTCTCGGAGAGACAATCACTGTACGAACATTCACCGAGGCTGATGAGCTTC 2123

## RESULT 3

US-09-103-840A-2/c  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: PRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Alignment Scores:  
Pred. No.: 84.2 Length: 4403765  
Score: 108.00 Matches: 46  
Percent Similarity: 35.42% Conservatives: 22  
Best Local Similarity: 23.96% Mismatches: 52  
Query Match: 10.39% Indels: 72

```

DB: 3 Gaps: 8
US-10-691-383-2_COPY_435_632 (1-198) x US-09-103-840A-1 (1-4411529)
QY 4 GlyThrSerHisTyr-----PheArgLeuIleGlyAla 14
D 4262534 GGAATCTCCACACTTCGGCGGAGCACTGATCGTTGGCTGATCTCGGCGCTGCTGGGCGC 4262475
QY 15 AlaGluLeuAlaGlnArgAlaSerCysTyrGlnIlystrpGlnValHisArgPheAlaArg 34
D 4262474 ATAGCCCTGCCACGGCGCCG-----CGGGAATGCTTGTG-----4262439
QY 35 ProGluAlaLeuGlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPhe 54
D 4262438 ---GCGGGGGCGGCGCTTCGTCGCCCATGCCATGCC-----4262403
QY 55 AspIleSerLeuLeuGluAsnAspGluLeuLeuIleArgValAlaGluIleAsnAlaAla 74
D 4262402 -----GTGCTGATCAAGCGGCTGTGTGCGG-----CGT 4262376
QY 75 GlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySerPro 94
D 4262375 CAGCGGCGGATCATCGGCCATC-----GCGGTCAACGTGGACACGCCA 4262331
QY 95 ThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeu 114
D 4262330 AGTCAACTGAGCTTTCGTCGGCACACGCCACCTCGACACGCGCGGCGCTGCTCATG 4262271
QY 115 LysAlaLeuIleGlyLeuAspArgGlyGluCysPheProAsnProValPheProSer 134
D 4262270 GGCAGACCCACGGGCTG-----CGGTACCGGTGTG-----4262238
QY 135 AspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIle 154
D 4262238 -----4262238
QY 155 AsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPhe 174
D 4262237 -----CTGTGCGCGCGATGCTGCGGAATACTGTGGGGGTCCACTACCCAGT 4262184
QY 175 AspGlyIleGlnGlyLeuLeuGlyGluThrIle 186
D 4262183 GATGTGCCGTGGGTGTGCTGTGGCGCCACCGTC 4262148

RESULT 4
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
Pred. No.: 84.5 Length: 4411529
Score: 108.00 Matches: 46
Percent Similarity: 35.42% Conservative: 22
Best Local Similarity: 23.96% Mismatches: 52
Query Match: 10.39% Indels: 72

DB: 3 Gaps: 8
US-10-691-383-2_COPY_435_632 (1-198) x US-09-103-840A-1 (1-4411529)
QY 4 GlyThrSerHisTyr-----PheArgLeuIleGlyAla 14
D 4270277 GGAATCTCCACACTTCGGCGGAGCACTGATCGTTGGCTGATCTCGGCGCTGCTGGGCGC 4270218
QY 15 AlaGluLeuAlaGlnArgAlaSerCysTyrGlnIlystrpGlnValHisArgPheAlaArg 34
D 4270217 ATAGCCCTGCCACGGCGCCG-----CGGGAATGCTTGTG-----4270182
QY 35 ProGluAlaLeuGlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPhe 54
D 4270181 ---GCGGGGGCGGCGCTTCGTCGCCCATGCCATGCC-----4270146
QY 55 AspIleSerLeuLeuGluAsnAspGluLeuLeuIleArgValAlaGluIleAsnAlaAla 74
D 4270145 -----GTGCTGATCAAGCGGCTGTGTGCGG-----CGT 4270119
QY 75 GlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySerPro 94
D 4270118 CAGCGGCGGATCATCGGCCATC-----GCGGTCAACGTGGACACGCCA 4270074
QY 95 ThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeu 114
D 4270073 AGTCAACTGAGCTTTCGTCGGCACACGCCACCTCGACACGCGCGGCGCTGCTCATG 4270014
QY 115 LysAlaLeuIleGlyLeuAspArgGlyGluCysPheProAsnProValPheProSer 134
D 4270013 GGCAGACCCACGGGCTG-----CGGTACCGGTGTG-----4269981
QY 135 AspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIle 154
D 4269981 -----4269981
QY 155 AsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPhe 174
D 4269980 -----CTGTGCGCGCGATGCTGCGGAATACTGTGGGGGTCCACTACCCAGT 4269927
QY 175 AspGlyIleGlnGlyLeuLeuGlyGluThrIle 186
D 4269926 GATGTGCCGTGGGTGTGCTGTGGCGCCACCGTC 4269891

RESULT 5
US-09-151-189-3
; Sequence 3, Application US/09151189
; Patent No. 6232457
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: Ng, Kwan L.
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses
; FILE REFERENCE: 023070-087100US
; CURRENT APPLICATION NUMBER: US/09/151,189
; CURRENT FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 51
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:probe for
; OTHER INFORMATION: second conserved region between Curvularia and
; OTHER INFORMATION: Ascopyllum vanadium peroxidase active sites
US-09-151-189-3

Alignment Scores:
Pred. No.: 2,79e-05 Length: 51
Score: 99.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

```

Query Match: 9.53% Indels: 0  
DB: 3 Gaps: 0

US-10-691-383-2\_COPY\_435\_632 (1-198) x US-09-151-189-3 (1-51)

QY 94 ProThHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPhe 110  
|||||  
DB 1 CCAACGACCCCTTCGTACCGCTGCGCACGCTACCCAAACGGAGCATTT 51

## RESULT 6

US-09-596-794-3  
; Sequence 3, Application US/09596794  
; Patent No. 6656715  
; GENERAL INFORMATION:  
; APPLICANT: Vreeland, Valerie  
; TITLE OF INVENTION: The Regents of the University of California  
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases  
; TITLE OF INVENTION: and Their Uses  
; FILE REFERENCE: 023070-087110US  
; CURRENT APPLICATION NUMBER: US/09/596,794  
; CURRENT FILING DATE: 2000-06-19  
; PRIOR APPLICATION NUMBER: US 09/151,189  
; PRIOR FILING DATE: 1998-09-10  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 51  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: probe for  
; OTHER INFORMATION: second conserved region between Curvularia and  
; OTHER INFORMATION: Ascopyllum vanadium peroxidase active sites  
US-09-596-794-3

Alignment Scores:  
Pred. No.: 2,79e-05 Length: 51  
Score: 99.00 Matches: 17  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 9.53% Indels: 0  
DB: 4 Gaps: 0

US-10-691-383-2\_COPY\_435\_632 (1-198) x US-09-596-794-3 (1-51)

QY 94 ProThHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPhe 110  
|||||  
DB 1 CCAACGACCCCTTCGTACCGCTGCGCACGCTACCCAAACGGAGCATTT 51

## RESULT 7

US-09-732-615-3  
; Sequence 3, Application US/09732615  
; Patent No. 6632650  
; GENERAL INFORMATION:  
; APPLICANT: Chen, Mario W.  
; APPLICANT: Chen, Qiong  
; APPLICANT: Gibson, Katherine J.  
; APPLICANT: Kostichka, Kristy N.  
; APPLICANT: Thomas, Stuart M.  
; APPLICANT: Nagarajan, Vasantha  
; TITLE OF INVENTION: Genes Involved in Cyclododecanone Degradation Pathway  
; FILE REFERENCE: BC1023 US NA  
; CURRENT APPLICATION NUMBER: US/09/732,615  
; CURRENT FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/170,214  
; PRIOR FILING DATE: December 10, 1999  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 3  
; LENGTH: 1161  
; TYPE: DNA  
; ORGANISM: Rhodococcus ruber  
; FEATURE:

US-09-732-615-3

Alignment Scores:  
Pred. No.: 0.0345 Length: 1161  
Score: 91.50 Matches: 50  
Percent Similarity: 39.66% Conservative: 21  
Best Local Similarity: 27.93% Mismatches: 74  
Query Match: 8.81% Indels: 34  
DB: 4 Gaps: 7

US-10-691-383-2\_COPY\_435\_632 (1-198) x US-09-732-615-3 (1-1161)

QY 14 AlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAla 33  
|||||  
DB 453 GCCGGCGGTTCCGGCCCGCGCTACCCACGCCAACACCGGGCCACTCATCGAGAAAT 512  
QY 34 ArgPro-----GluAlaLeuGlyThrLeuHisAsnThr-IleAlaGlyAs 49  
|||||  
DB 513 CGTCCGCGCATCACCGCGCGCGCTTCAAGGAGTTCTGCACACACATCGCGCGTCC 572  
QY 49 pLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuLeuLeuValAl 69  
|||||  
DB 573 GCTCGTCCGACTTCAGATCGCGCGCGCGAGAACGAT-----TGGGGCGCGTACCGC 626  
QY 69 aGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAla 89  
|||||  
DB 627 CGAGATCGTCGCACACACCCCTTCACATCGATCTCGCGCATCGATCCCGACAGCGT 686  
QY 89 eGlnValGlySerProThrHisPro-----SerTyrProSer-- 101  
: |||||  
DB 687 CATGTGAAGACGATGACGGGACCGGTAGCCGCGCGAACCGGCCAATACTCCCGGATG 746  
QY 102 -----GlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAl 116  
|||||  
DB 747 GCGGCACGCGCATCGGAGCCCTCAACGCTACGCGAACGCGACGCTCGCTCGCTCGGAT 806  
QY 116 aLeuIleGlyLeuAspArgGlyGlyCysPheProAsnProValPheProSerAspAs 136  
|||||  
DB 807 CTTCTCGACCATCACGTCGGCGGTGAG-----AGCAA 839  
QY 136 pGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIleAsn-- 155  
: |||||  
DB 840 CGGCTCTCGCTCTCTCCACAGCACCTCGATCGGCAAGATTTTCGAAGAGCAGAACACGA 899  
QY 156 ----LysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleHisTyr 172  
: |||||  
DB 900 CGTCGACCTCGTCTCGGTGTCGCGTCCGCGC-----GGCATCGGCTAC 945

## RESULT 8

US-09-732-615-13  
; Sequence 13, Application US/09732615  
; Patent No. 6632650  
; GENERAL INFORMATION:  
; APPLICANT: Chen, Mario W.  
; APPLICANT: Chen, Qiong  
; APPLICANT: Gibson, Katherine J.  
; APPLICANT: Kostichka, Kristy N.  
; APPLICANT: Thomas, Stuart M.  
; APPLICANT: Nagarajan, Vasantha  
; TITLE OF INVENTION: Genes Involved in Cyclododecanone Degradation Pathway  
; FILE REFERENCE: BC1023 US NA  
; CURRENT APPLICATION NUMBER: US/09/732,615  
; CURRENT FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/170,214  
; PRIOR FILING DATE: December 10, 1999  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 13  
; LENGTH: 10480  
; TYPE: DNA  
; ORGANISM: Rhodococcus ruber  
US-09-732-615-13



Alignment Scores:  
Pred. No.: 1.05 Length: 10480  
Score: 91.50 Matches: 50  
Percent Similarity: 39.66% Conservative: 21  
Best Local Similarity: 27.93% Mismatches: 74  
Query Match: 8.81% Indels: 34  
DB: 4 Gaps: 7

US-10-691-383-2\_COPY\_435\_632 (1-198) x US-09-732-615-13 (1-10480)

Qy 14 AlaAlaGluLeuAlaGlnAraGlnAraSerCysTyrGlnLysTrpGlnValHisArgPheAla 33  
Db 2951 GCCGGGGTTCGGCGCGGCTACACGCGCAACACGAGGCGCCATCTCATGAGAAAT 3010

Qy 34 ArgPro-----GluAlaLeuGlyThrLeuHisAsnThr-IleAlaGlyAs 49  
Db 3011 CGTCCGCGCATCACCGCGCGCGGTTCAAGAGTTCGTACACACACATCGCGGTCC 3070

Qy 49 pLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuLysArgValAl 69  
Db 3071 GCTCGCTCGCGACTTCAGATCGCGCGCGCGAGACGAT-----TGGGGCGGTACCGC 3124

Qy 69 aGluLeuAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaI 89  
Db 3125 CGAGATCGTGCACACCCCGCTTCGACATCGATCTCGCGCATCGATCCGACACGCT 3184

Qy 89 eGlnValGlySerProThrHisPro-----SerTyrProSer-- 101  
Db 3185 CATGGTGAACACGATGACGGGCGGTAGCGCGCGCAACGCGGCGCAATACTCCGGATG 3244

Qy 102 -----GlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAl 116  
Db 3245 CGGCGACCGCGACATGGGAGCCCTCAACGGTCAACGCGCAACGCGCTCGCTCGGAT 3304

Qy 116 aLeuLeuGlyLeuAspArgGlyGlyCysPheProAsnProValPheProSerAspAs 136  
Db 3305 CCTCTCGACCATCACGCTGGCGGTGAG-----AGCAA 3337

Qy 136 pGlyLeuGluLeuLeuAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluLeuAsn-- 155  
Db 3338 CGGCTCGGCTCTCTCACAGCACCTCGATCGGCAAGATTTTGAAGACGACGACACGA 3397

Qy 156 ----LysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyLeuHisTyr 172  
Db 3398 CGTCGACCTGCTCGTCTCGTGTGCGGTTCGCGCG-----GGCATCGGCTAC 3443

RESULT 9  
US-09-832-496-1  
; Sequence 1, Application US/09832496  
; Patent No. 6503508  
; GENERAL INFORMATION:  
; APPLICANT: Danielssen, Steffen  
; APPLICANT: Schneider, Palle  
; TITLE OF INVENTION: Polypeptides having haloperoxidase activity  
; FILE REFERENCE: 10042.200-US  
; CURRENT APPLICATION NUMBER: US/09/832.496  
; CURRENT FILING DATE: 2001-04-11  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 1815  
; TYPE: DNA  
; ORGANISM: Dreschlara hartleibii  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1815)  
US-09-832-496-1

Alignment Scores:  
Pred. No.: 0.231 Length: 1815  
Score: 87.50 Matches: 32  
Percent Similarity: 35.51% Conservative: 17  
Best Local Similarity: 23.19% Mismatches: 38  
Query Match: 8.42% Indels: 51  
DB: 4 Gaps: 5

US-10-691-383-2\_COPY\_435\_632 (1-198) x US-09-832-616-1 (1-1815)

Qy 71 IleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGln 90  
Db 1126 CTCGGCGCGCGCGCCACAAACACAAACAGACATCCCATTC----- 1164

Qy 91 ValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPhe 110  
Db 1165 -----AAGCGCGCATTCCTCCGCTACCCCTCTGTGTACGCGACCTTTGGCGGCGCGGTG 1218

Qy 110 ----- 110  
Db 1219 TTCMAATGTGCGTAGATACTACACGCGCGGTAGGAACATGGAACAGCAGCAGGCCA 1278

Qy 111 -----AlaThrValLeuLysAlaLeuLeuGlyLeuAspArgGlyGly 124  
Db 1279 GACAACATTGCCATCGACATGGTAGTATCCGAGGAGCTAAACCGTCTAAGCCGTGACCTA 1338

Qy 125 GluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeuLeu----- 141  
Db 1339 CGCAACGCTACGATCCACCGCGCCCAATCGAAGACGAGCCCGGTATCGTACGACCCGC 1398

Qy 142 -----AsnPheGluGlyAlaCys---LeuThrTyrGluGlyGluLeuAsnLysLeu 157  
Db 1399 GTCGTGCGACACTTCAACTCGGCTGGGAATTTATGTTGCAAAACGCAATTTACGAATT 1458

Qy 158 AlaValAsnValAlaPheGlyArgGlnMetLeuGlyLeuHisTyrArgPheAsp 175  
Db 1459 -----TTCCTGGGTGTCACACTGCGCTTTTCGAT 1485

RESULT 10  
US-09-832-616-1  
; Sequence 1, Application US/09832616  
; Patent No. 6506586  
; GENERAL INFORMATION:  
; APPLICANT: Danielssen, Steffen  
; APPLICANT: Schneider, Palle  
; TITLE OF INVENTION: Nucleic acids encoding polypeptides having haloperoxidase activity  
; FILE REFERENCE: 10175.200-US  
; CURRENT APPLICATION NUMBER: US/09/832.616  
; CURRENT FILING DATE: 2001-04-11  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 1815  
; TYPE: DNA  
; ORGANISM: Dreschlara hartleibii  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1815)  
US-09-832-616-1

Alignment Scores:  
Pred. No.: 0.231 Length: 1815  
Score: 87.50 Matches: 32  
Percent Similarity: 35.51% Conservative: 17  
Best Local Similarity: 23.19% Mismatches: 38  
Query Match: 8.42% Indels: 51  
DB: 4 Gaps: 5

US-10-691-383-2\_COPY\_435\_632 (1-198) x US-09-832-616-1 (1-1815)

Qy 71 IleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGln 90  
Db 1126 CTCGGCGCGCGCGCCACAAACACAAACAGACATCCCATTC----- 1164

Qy 91 ValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPhe 110  
Db 1165 -----AAGCGCGCATTCCTCCGCTACCCCTCTGTGTACGCGACCTTTGGCGGCGCGGTG 1218



```

Qy 112 ThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCysPhe----- 127
Db 1453 GAAC TGGCGAAACACACCTTGTAGTTGTAGGACAGGTCGCTGTTGTTGTTTCATCTGTAC 1512
Qy 128 -----ProAsnProValPheProSerAspGlyLeuGluLeu 140
Db 1513 AGTCCCGAAAGAATAAGAACCGCGGACCGCGCTCGAGGGGGAGACGGGGCGCCG--- 1569
Qy 141 IleAsnPheGluGlyAlaCysLeuThrTyrgluGlyGlu---IleAsnLysLeuAlaVal 159
Db 1570 -----CCCGGAGCTCAATCTAAGGGCTGGGGCGGGGTAAACACAGCTTCAGGTC 1620
Qy 160 AsnValAlaPheGlyArgGlnMet 167
Db 1621 GATATGAATTTGGGTATCAGATG 1644

RESULT 14
US-09-180-109A-25
; Sequence 25, Application US/09180109A
; Patent No. 6410293
; GENERAL INFORMATION:
; APPLICANT: MUKUMOTO, Fujio
; APPLICANT: NISHIO, Shoichi
; APPLICANT: AKIMARU, Jiro
; APPLICANT: MITSUDA, Satoshi
; TITLE OF INVENTION: DNA Fragments Containing Biotin Biosynthesase Gene and
; FILE OF INVENTION: Use of the Same
; FILE REFERENCE: 0152-0490P
; CURRENT APPLICATION NUMBER: US/09/180,109A
; CURRENT FILING DATE: 1998-12-03
; PRIOR APPLICATION NUMBER: 09/047838 JAPAN
; PRIOR FILING DATE: 1997-03-03
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 1358
; TYPE: DNA
; ORGANISM: Sphingomonas sp.
; FEATURE:
; OTHER INFORMATION: Strain = SC42405
; NAME/KEY: CDS
; LOCATION: (152)..(1207)
US-09-180-109A-25

Alignment Scores:
Pred. No.: 0.366 Length: 1358
Score: 84.50 Matches: 57
Percent Similarity: 37.18% Conservative: 30
Best Local Similarity: 24.36% Mismatches: 80
Query Match: 8.13% Indels: 67
DB: 4 Gaps: 13

US-10-691-383-2_COPY_435_632 (1-198) x US-09-180-109A-25 (1-1358)
Qy 12 IleGlyAlaAlaGluLeuAlaGlnArgAlaSerCys---TyrGlnLysTrpGlnVal--- 29
Db 59 ATAAGGATGCCACGCTGGATCGACGGCGGCTTGCCCTATGACCGGGGTTTGGCGCG 118
Qy 30 ---HisArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAla--- 47
Db 119 CGTCATCCCGCGGCAGACCGCGCGCTGAGGAATGACTATGACTGACACCCCGCCATC 178
Qy 48 -----GlyAspLeuAspAlaAspPheAspIleSerLeuLeu 59
Db 179 ACTGCAGTACCGACTGGACCGGTGAGGAATCGCGCGCTGTTCGACCTCGCGTTC--- 235
Qy 60 GluAsnAspGluLeuLeuLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsn 79
Db 236 -----ACCGAATGGTGTTCGCGGACGCCAAGTCCATCGCGCCACCATCCGACAC 289
Qy 80 GluValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyr 99

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 19, 2004, 03:58:58 ; Search time 2478.17 Seconds  
(without alignments)  
3463.005 Million cell updates/sec

Title: US-10-691-383-2\_COPY\_435\_632

Perfect score: 1039

Sequence: 1 VNFGTSHYFLRIGAAELAQ.....GLLGETITVRLHQLMTF 198

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US10691383/runat\_17092004\_102656\_1741/app\_query.fasta\_1.1230  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=BLOSUM62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORW=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USR=US10691383 @CGN 1 1 6663 @runat\_17092004\_102656\_1741 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.em.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.ey.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1039	100.0	2931	6	AR152190	Sequence
2	1039	100.0	2931	6	BD235853	Gene-modi
3	1039	100.0	2931	6	AR435585	Sequence
4	1039	100.0	2931	8	AF053411	Fucus dis
5	503.5	48.5	3420	8	LDI491787	Laminaria
6	502.5	48.4	3379	8	LDI491786	Laminaria
7	332.5	32.0	1797	6	AX201842	Sequence
8	332.5	32.0	1908	8	AF218810	Corallina
9	331.5	31.9	2035	8	D87657	Corallina p
10	329.5	31.7	1794	6	E17199	Corallina p
11	322	31.0	1791	6	E17200	Corallina p
12	322	31.0	2029	8	D87658	Corallina p
13	116.5	11.2	12198	1	AE001953	Sequence
C 14	108	10.4	17205	1	AE007185	Myobacte
C 15	108	10.4	244800	1	BX842584	Myobacte
C 16	108	10.4	278492	1	BX248347	Myobacte
C 17	104	10.0	300450	1	AP005960	Bradyrhiz
18	101.5	9.8	11205	1	AE013888	Yersinia
C 19	101.5	9.8	199050	1	AJ414147	Yersinia
20	101	9.7	5828	6	E47157	Method for
21	99	9.5	51	6	AR152191	Sequence
22	99	9.5	51	6	BD235854	Gene-modi
23	99	9.5	51	6	AR435586	Sequence
24	99	9.5	302550	1	EX294137	Pirellula
C 25	97	9.3	302178	1	AE016918	Chromobac
C 26	93	9.0	325483	1	AP005050	Streptomy
27	92.5	8.9	1458	8	AY045606	Arabidops
28	91.5	8.8	1161	6	AR408895	Sequence
29	91.5	8.8	1161	6	AX167387	Sequence
30	91.5	8.8	10480	1	AY052630	Rhodococc
31	91.5	8.8	10480	6	AR408900	Sequence
32	91.5	8.8	10480	6	AX167397	Sequence
33	90.5	8.7	1389	8	AY139975	Arabidops
34	90.5	8.7	3107	5	BC054581	Danio rer
35	90.5	8.7	172289	2	AC146249	Pan trogl
36	89	8.6	37977	9	AC112718	Homo sapi
37	89	8.6	169828	9	AC097461	Homo sapi
38	88.5	8.5	1944	6	AX191492	Sequence
39	88.5	8.5	2117	6	AX083511	Sequence
40	88.5	8.5	2122	6	AX191502	Sequence
41	88.5	8.5	2178	9	BC062625	Homo sapi
42	88.5	8.5	2192	9	AK025537	Homo sapi
43	88.5	8.5	2248	9	AK000695	Homo sapi
44	88.5	8.5	3987	1	SAC550389	Sulfolobu
C 45	88.5	8.5	24985	1	AY372755	Bacteroid

ALIGNMENTS

RESULT 1

```
AR152190
LOCUS AR152190 2931 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 1 from patent US 6232457.
ACCESSION AR152190
VERSION AR152190.1 GI:15118240
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
  1 (bases 1 to 2931)
AUTHORS Vreeland V. and Ng K.L.
TITLE Recombinant vanadium haloperoxidases and their uses
JOURNAL Patent: US 6232457-A 1 15-MAY-2001;
FEATURES
  Location/Qualifiers
    source
      1..2931
      /organism="unknown"
      /mol_type="unassigned DNA"
ORIGIN
Alignment Scores: 1.26e-92 Length: 2931
Pred. No.: 1039.00 Matches: 198
Score: 1039.00
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-691-383-2_COPY_435_632 (1-198) x AR152190 (1-2931)
QY 1 ValAsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArg 20
DB 1530 GTCAACTTCGGCAGCTCTCACTTTCAGATTGATAGTCCCGCGGAGCTGGCGCAGCGT 1589
QY 21 AlaserCysTyrGlnIleValHisArgPheAlaArgProGluAlaLeuGlyGly 40
DB 1590 GCCTCGTGTACCAAAAGTGGCAGGTGCATCGATTTCACGCCCGCGAGGCTCTCGGGGT 1649
QY 41 ThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGlu 60
DB 1650 ACCCTCCACAAACACCATCGCGGGGATCTAGATGCAGATTCGACATCTCCCTTCTTGA 1709
QY 61 AsnAspGluLeuLeuIleValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGlu 80
DB 1710 AATGATGAGCTCTTGAACGTGTGGCGAGATAAATCGCGCGAGATCCCAACACGAG 1769
QY 81 ValThrTyrLeuLeuProGlnAlaIleGlnValIleGlySerProThrHisProSerTyrPro 100
DB 1770 GTCACTTACCTTCTTCCACAGCTATCCAAAGTGGATTCGACGCCCGCTTCTTACCCG 1829
QY 101 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuIleGlyLeu 120
DB 1830 TCCGGCCACGCTACCCAAATGGAGCATTTGCCACAGTTCTGAAGGCCCTCATTTGGCCTA 1889
QY 121 AspArgGlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeu 140
DB 1890 GATCGGGAGGTGAGTCTTCCCTAACCCCGTGTCCAGCGATGACGCGCTTGGAACTA 1949
QY 141 IleAsnPheGluGlyAlaCysLeuThrTyrGlnIleGlyLeuIleGlyLeu 160
DB 1950 ATCAACTTCGAAGGGGATGCTTACATATGAGGAGAGATCAACAGCTCGCGGTCAAC 2009
QY 161 ValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeu 180
DB 2010 GTCGCATTTGGGAGGAGATGCTGGGCATTCACATATCGGTTCCAGCGATTCGAAGGCCCTA 2069
QY 181 LeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThrPhe 198
DB 2070 CTTCGAGAGACAACTCACTGTACGAACATTCACCGAGGAGCTGATGACGTTTC 2123
RESULT 2
BD235853
LOCUS BD235853 2931 bp DNA linear PAT 17-JUL-2003
DEFINITION Gene-modified vanadium haloperoxidase and utilization thereof.
```

```
BD235853
ACCESSION BD235853.1 GI:33045623
VERSION BD235853.1
KEYWORDS JP 2002525046-A/1.
SOURCE Fucus gardneri
ORGANISM Fucus gardneri
REFERENCE
  1 (bases 1 to 2931)
AUTHORS Vreeland V. and Ng K.L.
TITLE Gene-modified vanadium haloperoxidase and utilization thereof
JOURNAL Patent: JP 2002525046-A 1 13-AUG-2002;
  THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
COMMENT
  OS Fucus gardneri
  PN JP 2002525046-A/1
  PD 13-AUG-2002
  PF 27-AUG-1999 JP 2000570298
  PR 10-SEP-1998 US 09/151189
  PI VALERIE VREELAND, KWAN L NG
  PC C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12N9/02, C12N15/
  PC 00, C12N5/00
  CC Gene-modified vanadium haloperoxidase and utilization thereof
  FH Key Location/Qualifiers
  FT CDS
FEATURES
  source
    1..2931
    /organism="Fucus gardneri"
    /mol_type="genomic DNA"
    /db_xref="taxon:3013"
ORIGIN
Alignment Scores: 1.26e-92 Length: 2931
Pred. No.: 1039.00 Matches: 198
Score: 1039.00
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-691-383-2_COPY_435_632 (1-198) x BD235853 (1-2931)
QY 1 ValAsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArg 20
DB 1530 GTCAACTTCGGCAGCTCTCACTTTCAGATTGATAGTCCCGCGGAGCTGGCGCAGCGT 1589
QY 21 AlaserCysTyrGlnIleValHisArgPheAlaArgProGluAlaLeuGlyGly 40
DB 1590 GCCTCGTGTACCAAAAGTGGCAGGTGCATCGATTTCACGCCCGCGAGGCTCTCGGGGT 1649
QY 41 ThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGlu 60
DB 1650 ACCCTCCACAAACACCATCGCGGGGATCTAGATGCAGATTCGACATCTCCCTTCTTGA 1709
QY 61 AsnAspGluLeuLeuIleValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGlu 80
DB 1710 AATGATGAGCTCTTGAACGTGTGGCGAGATAAATCGCGCGAGATCCCAACACGAG 1769
QY 81 ValThrTyrLeuLeuProGlnAlaIleGlnValIleGlySerProThrHisProSerTyrPro 100
DB 1770 GTCACTTACCTTCTTCCACAGCTATCCAAAGTGGATTCGACGCCCGCTTCTTACCCG 1829
QY 101 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuIleGlyLeu 120
DB 1830 TCCGGCCACGCTACCCAAATGGAGCATTTGCCACAGTTCTGAAGGCCCTCATTTGGCCTA 1889
QY 121 AspArgGlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeu 140
DB 1890 GATCGGGAGGTGAGTCTTCCCTAACCCCGTGTCCAGCGATGACGCGCTTGGAACTA 1949
QY 141 IleAsnPheGluGlyAlaCysLeuThrTyrGlnIleGlyLeuIleGlyLeu 160
DB 1950 ATCAACTTCGAAGGGGATGCTTACATATGAGGAGAGATCAACAGCTCGCGGTCAAC 2009
QY 161 ValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeu 180
```

Db 2010 GTCCGATTTCGGAGGACGATGCTGGGCGATCCACTATCGGTTCCGACGGTATCCAAAGGCGCTA 2069

Qy 181 LeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThrPhe 198  
|||||  
Db 2070 CTCTCGGAGAGACAATCACTGTACGAACACTTCCACGAGGCTGATGACGTTTC 2123

RESULT 4  
AF053411

LOCUS AF053411 2931 bp DNA linear PAT 18-DEC-2003

DEFINITION Sequence 1 from patent US 6656715.

ACCESSION AR435585

VERSION AR435585.1 GI:40198553

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 2931)

AUTHORS Vreeland,V. and Epstein,L.

TITLE Recombinant minimal catalytic vanadium haloperoxidases and their uses

JOURNAL Patent: US 6656715-A 1 02-DEC-2003;

FEATURES  
Location/Qualifiers  
1..2931  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN

Alignment Scores:  
Pred. No.: 1,26e-92 Length: 2931  
Score: 1039.00 Matches: 198  
Percent Similarity: 100.00% Conservations: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-10-691-383-2\_COPY\_435\_632 (1-198) x AR435585 (1-2931)

Qy 1 ValAsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArg 20  
|||||  
Db 1530 GTCAATTCGGCAGCTCTCACTACTTCCAGATTGATAGTCCCGCGAGCTGGCGCAGCGT 1589

Qy 21 AlAserCysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGly 40  
|||||  
Db 1590 GCCTCGGTATACCAAAAGTGGCAGGTGCATGATTTCCACGCCCGCGAGGCTCTCGGGGT 1649

Qy 41 ThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGlu 60  
|||||  
Db 1650 ACCCTCCACACACCATCGCGGGGATCTAGATGCAGACTTCGACATCTCCCTTCTTGA 1709

Qy 61 AsnAspGluLeuLeuLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGlu 80  
|||||  
Db 1710 AATGATCAGCTCTTGAACAGTGTGGCGAGATAAATCGCGCAGAGATCCCAACACGAG 1769

Qy 81 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 100  
|||||  
Db 1770 GTCACTTACCTTCTTCACACAGCTATCCAAAGTGGATCGCAACGCCACCTTCCTACCCG 1829

Qy 101 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeu 120  
|||||  
Db 1830 TCCGGCCACGCTACCCAAATGGAGCAATTTGCCACAGTTCTGAAGGCCCTCATTTGGCTA 1889

Qy 121 AspArgGlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeu 140  
|||||  
Db 1890 GATCGGGAGGTGAGTGCTTCCCTAACCCCGTGTTCCCAAGCGATGACGCGCTGGAAC 1949

Qy 141 IleAsnPheGluGlyAlaCysLeuThrTyrGluGlyIleAsnLysLeuAlaValAsn 160  
|||||  
Db 1950 ATCAACTTCGAAGGGGATGCTTACATATGAGGAGAGATCAACAAGCTCGCGGTCAAC 2009

Qy 161 ValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeu 180  
|||||  
Db 2010 GTCCGATTTCGGAGGACGATGCTGGGCTATCCACTATCGGTTCCGACGGTATCCAAAGGCGCTA 2069

Qy 181 LeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThrPhe 198  
|||||  
Db 2070 CTCTCGGAGAGACAATCACTGTACGAACACTTCCACGAGGCTGATGACGTTTC 2123

RESULT 4  
AF053411

LOCUS AF053411 2931 bp mRNA linear PLN 12-SEP-1998

DEFINITION Fucus distichus vanadium bromoperoxidase mRNA, complete cds.

ACCESSION AF053411

VERSION AF053411.1 GI:3582762

KEYWORDS

SOURCE Fucus distichus

ORGANISM Fucus distichus

REFERENCE 1 (bases 1 to 2931)

AUTHORS Vreeland,V., Ng,K. and Epstein,L.

TITLE Direct Submission

JOURNAL Submitted (11-MAR-1998) Environmental Science, Policy and Management, University of California, 201 Wellman Hall, Berkeley, CA 94720-3112, USA

FEATURES  
Location/Qualifiers  
1..2931  
/organism="Fucus distichus"  
/mol\_type="mRNA"  
/db\_xref="taxon:3012"  
/dev\_stage="2-cell embryo"  
1..227  
228..2258  
/codon\_start=1  
/product="vanadium bromoperoxidase"  
/protein\_id="AAC35279.1"  
/db\_xref="GI:3582763"  
/translations="MLCHAADTTRGSPMDPTGVRLTTSQRKGMRRQLEGEKSLGF  
HPSETPVKYLEGSETWKVKLTGDISASKILGKIMARVRIATATAVLAAPCLAFD  
EVTASGVPEEHKHTGEGHLOCTWNSDDALDTPNRDNVAFASRDAARRRDEGT  
GTVCQTNGETDLATMFHKSPLHDELGCQTADDFALIEDCLNGDFSI CEDVPAGDPA  
GRUVNTPAFAIDISGPASATTPVPTLSSPELAQAELVWMLARDVPFMQYGT  
DETTTAAANLAGMGFPNLDVSIIGSDGTVPFSQLFRATFVGVEVGPVSQLVNS  
FTDATTVPKQETFAFDNLNMYDFEMLNIONGGPPAGPEELDEELRFIRNARDLAR  
VSPVDINTEAYRGVSLILLELGAFSRPGINGPFIIDROAGFVNFVSHYFRLLTGAAE  
LAORASYCKWQVHRFARPEALGTLHTNTIAGDLADFDI SLENDELLAKRVAENAA  
QNPNEVYLLPOAIQVGSPTSPSGHATQNGAFATVLAALLGLDRGEGCEFPNVPF  
PSDGLHLEFEGACLTGEGINKLVANVAFGRQMLGIHRYFDIGIQLLGETITVIRT  
LHQLMTFAEEATFEFLFTGEVTKLFDQDFTFSDGDMCSGLVYTVTGADCOA"

5'UTR  
CDS

2259..2931

ORIGIN

Alignment Scores:  
Pred. No.: 1,26e-92 Length: 2931  
Score: 1039.00 Matches: 198  
Percent Similarity: 100.00% Conservations: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 8 Gaps: 0

US-10-691-383-2\_COPY\_435\_632 (1-198) x AF053411 (1-2931)

Qy 1 ValAsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArg 20  
|||||  
Db 1530 GTCAATTCGGCAGCTCTCACTACTTCCAGATTGATAGTCCCGCGAGCTGGCGCAGCGT 1589

Qy 21 AlAserCysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGly 40  
|||||  
Db 1590 GCCTCGGTATACCAAAAGTGGCAGGTGCATGATTTCCACGCCCGCGAGGCTCTCGGGGT 1649

Qy 41 ThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGlu 60  
|||||  
Db 1650 ACCCTCCACACACCATCGCGGGGATCTAGATGCAGACTTCGACATCTCCCTTCTTGA 1709

Qy 61 AsnAspGluLeuLeuLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGlu 80  
|||||  
Db 1710 AATGATCAGCTCTTGAACAGTGTGGCGAGATAAATCGCGCAGAGATCCCAACACGAG 1769

Qy 81 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 100  
|||||  
Db 1770 GTCACTTACCTTCTTCACACAGCTATCCAAAGTGGATCGCAACGCCACCTTCCTACCCG 1829

Qy 101 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeu 120  
|||||  
Db 1830 TCCGGCCACGCTACCCAAATGGAGCAATTTGCCACAGTTCTGAAGGCCCTCATTTGGCTA 1889

Qy 121 AspArgGlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeu 140  
|||||  
Db 1890 GATCGGGAGGTGAGTGCTTCCCTAACCCCGTGTTCCCAAGCGATGACGCGCTGGAAC 1949

Qy 141 IleAsnPheGluGlyAlaCysLeuThrTyrGluGlyIleAsnLysLeuAlaValAsn 160  
|||||  
Db 1950 ATCAACTTCGAAGGGGATGCTTACATATGAGGAGAGATCAACAAGCTCGCGGTCAAC 2009

Qy 161 ValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeu 180  
|||||  
Db 2010 GTCCGATTTCGGAGGACGATGCTGGGCTATCCACTATCGGTTCCGACGGTATCCAAAGGCGCTA 1769

```

Qy 81 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 100
Db 1770 GTACCTACCTCTTCCACAGCTATCCAGTGGATCGGCAACGACCTTCTTACCCG 1829
Qy 101 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeu 120
Db 1830 TCCGCCACCGCTACCCAAAATGGAGCATTTGCCACAGTTCTGAAGGCCCTCATTTGGCCTA 1889
Qy 121 AspArgGlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeu 140
Db 1890 GATCGGGAGGTGAGTCTTCCCTAACCCGCTGTTCCCAAGCGATGACGGCTTGGAACTA 1949
Qy 141 IleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsn 160
Db 1950 ATCAACTTCGAAGGGGATGCCCTTACATATGAGGAGAGATCAACAGCTCGCGGTCAAC 2009
Qy 161 ValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeu 180
Db 2010 GTCGCATTTGGAGGAGATGCTGGGCATCCATCTATCGTTCCAGCGGTATCCAAAGGCCTA 2069
Qy 181 LeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThrPhe 198
Db 2070 CTTCTCGAGAGACAATCACTGTACGAACACTTCCACAGAGCTGATGACGTTTC 2123

RESULT 5
LDI491787 3420 bp mRNA linear PLN 28-JUN-2003
LOCUS Laminaria digitata mRNA for vanadium-dependent bromoperoxidase 2
DEFINITION (VBPO2 gene).
ACCESSION AJ491787
VERSION AJ491787.1 GI:32329413
KEYWORDS vanadium-dependent bromoperoxidase 2; VBPO2 gene.
SOURCE Laminaria digitata
ORGANISM Laminaria digitata
            Eukaryota; stramenopiles; Phaeophyceae; Laminariales;
            Laminariaceae; Laminaria.
REFERENCE 1
AUTHORS Colin,C., Leblanc,C., Wagner,E., Delage,L., Leize-Wagner,E., Van
Dorselaer,A., Klorreg,B. and Potin,P.
TITLE The brown algal kelp Laminaria digitata features distinct
bromoperoxidase and iodo-peroxidase activities
J. Biol. Chem. 278 (26), 23545-23552 (2003)
JOURNAL 2703831
MEDLINE 22697758
PUBMED 2
(bases 1 to 3420)
REFERENCE 2
AUTHORS Leblanc,C.
TITLE Direct Submission
SUBMITTED (26-JUN-2002) Leblanc C., UMR 1931, CNRS-Goemar, Station
Biologique, BP 74, 29682 Roscoff, FRANCE
JOURNAL
FEATURES
source
1..3420
    /organism="Laminaria digitata"
    /mol_type="mRNA"
    /db_xref="taxon:80365"
    /tissue_type="sporophyte"
    /clone_lib="lambda ZAP II - S7"
    /country="France; Brittany"
1..3420
    /gene="VBPO2"
1..36
    /gene="VBPO2"
37..2085
    /gene="VBPO2"
    /codon_start=1
    /product="vanadium-dependent bromoperoxidase 2"
    /protein_id="CAD37192.1"
    /db_xref="GI:32329414"
    /translation="MKEARPTOPCALYSVLSLRHVACALCVACAFLLFEYQVSSLGL
TALLAAPACGLGVEEPEPTQPLLSGNCVRSRSLDFDPVPRAKVTLTKRLAIKDE
ISVGTCPNVNNGDEENPLFAGQYHKLPHDKFGQVDEDAYKLLCEVFSDFINECEK
VPSGARRGAKLTNPJGGTAHQVTGADSDNVFTTTPDSLLSRLLAAQAEVYWMALL
RDI1PFGFAKNDYVRLAAENLQSLPAFKGLN1PRSEGGKIDPVD1LFRITWPGVTTGP
VWSQFMLSDFLIDS1K1VTPKADPLTPGVDMYMTAFQPWLDVQNGASKLETTFDSENPFR

```

```

INRDLATIALRDOLYTEAFRAALILFTGALGGEVGPVAAERQCGFATFGFPHIL
TAWASSTRHAWYAKQVHMLRPEAYCALVHNTLMRDVITPLPDSILRNTELLNR
VEVHNRNMPDGEKIFLLPMAAAGSPTHPAPSGHAINNGAYITALKAFGLYEAQK
CFNPVSVNDEGTRIKYKPSREIVGECNEKGLUVEGLITYEGSLINKISANVLLGRS
HIGVHRMDGVYALMGSETSCVRRLOQELPGLPEAREVEGKKRRGDIPTATYKFLYS
GKILELYGNLYKLDGKLCGAGTGDDFCDPIDDEDYSSFDDIVESHAFOSLHHTL
"
2086..3420
    /gene="VBPO2"

Alignment Scores:
Pred. No.: 1,01e-39 Length: 3420
Score: 503.50 Matches: 98
Percent Similarity: 60.00% Conservatve: 28
Best Local Similarity: 46.67% Mismatches: 67
Query Match: 48.46% Indels: 17
DB: 8 Gaps: 2

US-10-691-383-2_COPY_435_632 (1-198) x LDI491787 (1-3420)
Qy 3 PheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSer 22
Db 1192 TTCCGGGAGCGGCACATCTCTCACCGCATGCGCTCGGCCAGCTCTTCCACGCGCACGCG 1251
Qy 23 CysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeuGlyThrLeu 42
Db 1252 TGTAGTCGTAAGTGGCAAGTCCACCGCATGCTCGACCGCGGAGCGTACGCTGCGCTGGTA 1311
Qy 43 HisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAsp 62
Db 1312 CATAACACGCTCATGAGGAGCGTCAACCCCGCTCGCGGACTCTATCTCTGAGGAACACG 1371
Qy 63 GluLeuLeuLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThr 82
Db 1372 GAGCTCTCTCAACCGAGTGGAGGTGCACCAACCGCATGTATGAACCGCACGCGCGAGAAGACC 1431
Qy 83 TyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGly 102
Db 1432 TTCTCTGCTCCCATGTCGGGAGCGCCAGCGCTCTCCACGACCCCGCTTACCCAGTGGGA 1491
Qy 103 HisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArg 122
Db 1492 CACGCCATCAACAACGCGCTACATCACCGCGCTCAAGCGCTTCTCGGCTACGAGGCC 1551
Qy 123 GlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeuIleAsn 142
Db 1552 GGCAGAAAGTGTCTTCCCAACCCCGTGTCTCGAACGACGAGGAACCAAGAGGATCAAA 1611
Qy 143 PheGlu-----GlyAlaCys----- 147
Db 1612 TACAAGCCCTCCGAAGGGAGATTGGGGAGTGTCTCAACGAGAGGCGCAAGCTCGTC 1671
Qy 148 -----LeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArg 165
Db 1672 GAAGGCTTCACGTACGAGGAGAGAGTCAACAAGATCAGCGCCCAACGCTCTTGGGAAGG 1731
Qy 166 GlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThr 185
Db 1732 TCCCATCATCGGTGTTCACTGGAGGATGGACGGCGTGTACGAGGCGCTCATGGGAGAGACC 1791
Qy 186 IleThrValArgThrLeuHisGlnGluLeu 195
Db 1792 AGCTGCTCCGCCCTGTCAGCAGGAACCTT 1821

RESULT 6
LDI491786 3379 bp mRNA linear PLN 28-JUN-2003
LOCUS Laminaria digitata mRNA for vanadium-dependent bromoperoxidase 1
DEFINITION (VBPO1 gene).
ACCESSION AJ491786
VERSION AJ491786.1 GI:32329411
KEYWORDS vanadium-dependent bromoperoxidase 1; VBPO1 gene.

```









```

Db      1120 GTGAACCTCGGATCCGACACGCTGCTGAGTCTGCTGACTGAAGTGGCTACGCGCGCTTG 1179
Qy      20  ArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeuGly 39
      ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1180 AAGCGCGGTACGGTACCAGAGTTTAAACATTTCATCGCTCGCTGCGCCCTGAGGCTACCGGT 1239
Qy      40  Gly---ThrLeuHisAsnThrIleAla-----GlyAspLeuAspAlaAspPhe 54
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1240 GGTCTGATTATCGTTTAAACAAATCGCACCCGAGAGGGCGAGAGCATTTCCCTGAGTT 1299
Qy      55  AspIleSerLeuLeuGluAsnAspGluLeuLeuLysArgValAlaGluIleAsnAlaAla 74
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1300 GATCTTCTGTTGAAGAGCTTGAGATATCTTGAGAAAGCTGAAATTAGCAATAGGAAG 1359
Qy      75  GluAsn-----ProAsnAsnGluValThrTyrLeuLeuProGlnAlaIle 89
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1360 CAGAACAATAGCTGACGAGATCTGACCCGTGATCCCTTCATTCCTGTTGCCGATGGCATTC 1419
Qy      90  GlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAla 109
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1420 GCCGAGGCGAGCCCATTCATCCGTCCTAGGAGCGGCCACCGCTGTGGTGTGCTGGCGCA 1479
Qy      110 PheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCysPheProAsn 129
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1480 TGTGTGACGATCTCGAAGCGGTTTCTC-----TTCGACTCCGCGCATCGAGATC-----GAT 1527
Qy      130 ProValPheProSerAsp-----AspGlyLeuGluLeuIleAsnPheGluGlyAla 146
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1528 CAGGTGTTCGAGGTGCGACAAAGATGAGGACAAGCTTGTGAAGTCGCTTTTCAAGGGGAAC 1587
Qy      147 CysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGln 166
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1588 ---CTCACTGTTCCGCGTGAATTGAACAAGCTCGCCGACATATTGCGATCGGCGGTAA 1644
Qy      167 MetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThrIle 186
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1645 ATGCGAGGTGTTCACTACTTCTCTGACCAAGTTCAGTTCAGTCTTCTGCTCGGTGAGCAGTT 1704
Qy      187 ThrValArgThrLeuHisGlnGluLeuMetThrPhe 198
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1705 GCGATTGGAATCTTGAAGAGCAAGTCTGACGTAT 1740

RESULT 11
LOCUS   E17200              1791 bp      DNA      linear      PAT 28-JUL-1999
DEFINITION Corallina pilulifera mRNA for haloperoxidase.
ACCESSION E17200
VERSION   E17200.1 GI:5711883
KEYWORDS  JP 1998248581-A/2.
SOURCE    unidentified
ORGANISM  unidentified
REFERENCE 1 (bases 1 to 1791)
AUTHORS   Izumi,Y. and Tanabe,T.
TITLE     NEW HALOPEROXIDASE GENE AND ITS UTILIZATION
JOURNAL   Patent: JP 1998248581-A 2 22-SEP-1998;
          OTSUKA PHARMACEUT CO LTD
COMMENT   OS Corallina pilulifera
          PN JP 1998248581-A/2
          PD 22-SEP-1998
          PF 06-MAR-1997 JP 1997070539
          PI IZUMI YOSHITAKU, TANABE TADASHI
          PC C12N15/09,C12N9/08,(C12N15/09,C12R1:89),(C12N9/08,C12R1:19);
          CC strandedness: Double;
          CH topology: Linear;
          FH Key Location/Qualifiers
          FI source 1..1791
          FT /clone='BPO2' /organism='Corallina pilulifera' FT
          FT mat_peptide 1..1791
          FT /product='haloperoxidase'.
          FT Location/Qualifiers
FEATURES

```

```

source 1..1791
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

ORIGIN
Alignment Scores:
Pred. No.: 3..78e-22 Length: 1791
Score: 322.00 Matches: 87
Percent Similarity: 52.97% Conservative: 29
Best Local Similarity: 39.73% Mismatches: 71
Query Match: 30.99% Indels: 32
DB: 6 Gaps: 8

US-10-691-383-2_COPY_435_632 (1-198) x E17200 (1-1791)
Qy      1  ValAsnPheGlyThrSerHisTyrPheArgLeuIle---GlyAlaAlaGluLeuAlaGln 19
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1114 GTGAACCTTCGATCCGACACACGCTGCTGAGTCTGCTGACTGAAGTGGCTACGCGCGCTTG 1173
Qy      20  ArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeuGly 39
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1174 AAGCGCGTACGGTATCAGAAAGTTTAAACATTTCATTCGCTGCGCGCTGAGGCTACTCGT 1233
Qy      40  GlyThrLeu-----HisAsnThrIleAlaGlyAspLeuAspAlaAspPheAsp 55
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1234 GGTCTGATTATCGTTTAAACAAGAAATCTTCTCTTGGCGGTAGT-----GAT 1278
Qy      56  IleSerLeuLeuGluAsnAspGluLeuLeuLysArg-----Val 68
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1279 ATAATATTCCCTGAAGTTAGTGAACCTGGTCGAGGAGCTCTCATCAATTCTGGATGACGTT 1338
Qy      69  AlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluVal-----ThrTyr 83
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1339 GCTGAGAGCAATGAAAACAGAACAGGCTGAGCGGATTTGAGCCCGGATAATCATATT 1398
Qy      84  LeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHis 103
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1399 CTGTTGCCGATGCAATTTGCCGAGGCGACGCCATTCATCCGCTCTATGGAAGTGGCCAC 1458
Qy      104 AlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGly----- 119
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1459 GCTGTGTTGCTGCGCATGTGTGCAATCTCAAGCGCTTCTTCGACGCCAATCTCCAG 1518
Qy      120 LeuAspArgGlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGlu 139
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1519 ATCGATAAGGTGTTTCGAG-----GTCGACACTGATGAGGACAGCTTGTG 1563
Qy      140 LeuIleAsnPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGly 159
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1564 AAGTCGTCITTTCAAGGGAAC---CTCACGTGTTGCGGTGAATTGAACAAGCTCGCCGAC 1620
Qy      160 AsnValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGly 179
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1621 AATGTTGCGATCGCGGGAACATGCGGGGTGTTCTACTACTTCTCTGACCAAGTTCGAGTCA 1680
Qy      180 LeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThrPhe 198
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1681 CTGCTACTCGGTGAACAGATTGCGATTGGAATCTTGGAGGAGCAGACCTTGACGTAT 1737

RESULT 12
LOCUS   D87658
DEFINITION Corallina pilulifera mRNA for vanadium-dependent bromoperoxidase 2, complete cds.
ACCESSION D87658
VERSION   D87658.1 GI:3293051
KEYWORDS  vanadium-dependent bromoperoxidase 2.
SOURCE    Corallina pilulifera
ORGANISM  Corallina pilulifera
          Eukaryota; Rhodophyta; Florideophyceae; Corallinales;
          Corallinaceae; Corallinoideae; Corallina.
          1 (bases 1 to 2029)

```



KDLTALRVASPGALTTAELLRLRAYPGVEVVRMYRDEVPVAVORGEYDGTAVDAGLII  
HESRTVPEYGLDKLLDGLAWGSETGLPLPGAILVRRDLPAQTORELODAVRRSLE  
YAYAPAAADYIREHALEMSDEVMQAHIDLYNLSLDVGEGERAVRELLRRAVEV  
GAVTARPELFFVGDAEAGTPTIL"

gene

1033..1956

CDS

/gene="DR1012"

1033..1956

/gene="DR1012"

/notes="similar to GB:AL009126 percent identity: 54.74;

identified by sequence similarity; putative"

/codon\_start=1

/transl\_table=11

/product="ABC transporter, ATP-binding protein"

/protein\_id="AAFI0588.1"

/db\_xref="GI:6458742"

/translation="MSKASAPAPPAVEVRGLSKSYGRANVLTDFLVNIPGEVYALTG  
PNAGKTTLRIMTIGAFPPRGVIRGVNNGVNDPPRAALMGAVPAKFPQFTG  
TENLKHASITAGVGPRIGSEVLALELTSMGRRVAEYSLGQORQLGVASAILAEP  
KVLIDPETSGLDPLGIGLTHRIVTSLATSGCAVILSTHLREIATVAHTVGTMTGR  
LVTDVLRSTSAVRFDVDDPVGAALLERQPYVRKATPRPPYAVAGLGAQANVPEAL  
RLANEDVAYVEYTPDHFIDLYEYRVEYA"

gene

1956..2729

CDS

/gene="DR1013"

1956..2729

/gene="DR1013"

/notes="identified by Glimmer2; putative"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAFI0594.1"

/db\_xref="GI:6458748"

/translation="MLSLIERKLLARSARIALVCTLLPPLMSFAPRLGEILRV  
DIVSGQLPAVSGVAVQFLEPLFIYAVAEIMCAEVAQCTLAPLILRPVRSRRIVS  
KLIVLFPPLLVLGLSLVAGIPHYGEFRGGTGLGFLQGVGLVSLPSAALAEV  
LRGLLTIWMLPVALSLFYGLLSTSAALATLATLSVRLVLLVFPFGLQORVLLT  
SYNLXYTQSLTQALVLLIYTLGFSLMAIFAPDRDL"

gene

complement(2935..3372)

CDS

/gene="DR1014"

complement(2935..3372)

/gene="DR1014"

/notes="similar to GP:188406 percent identity: 57.35;

identified by sequence similarity; putative"

/codon\_start=1

/transl\_table=11

/product="vanadium chloroperoxidase-related protein"

/protein\_id="AAFI0589.1"

/db\_xref="GI:6458743"

/translation="MTPLGLWIBEALRGQARLGSDLAQVLAATAVAGHDAFISCW  
QKGFYVNAVQSMWDHVPQGWAPSLPTPPPPSPVPSGHATVSGAAAEVLAQFPPLQAR  
QLRRDARDAFSAFVVGSIHNGVDGVLGVQVRVARALLEKP"

gene

complement(3423..4277)

CDS

/gene="DR1015"

complement(3423..4277)

/gene="DR1015"

/notes="identified by Glimmer2; putative"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAFI0595.1"

/db\_xref="GI:6458749"

/translation="MSRASLSHWQPHRRRLRWCAVSCVIGIAGAGAVAQSSAVOSL  
TAPSLPQTQAPALPAVRSGLAASPLVISPAPXLRQTRSAALAAAPAGPATALSVO  
VPRVARTIDLLIARLDLPPPARALATLGLHDAWLARQSGRAVDSDTAAAVAA  
TRQMAUFDRDASLSELSALPPLPAEVDRAASGLGAVADAVLAPAQDQAQNS  
QGNAAQSGSGWAPLSGQALEPFGMRVQPIGLAPDALPQSRPPRAGRAPSPPTAS  
VSPTSREP"

gene

4376..5707

CDS

/gene="DR1016"

4376..5707

/gene="DR1016"

/notes="similar to SP:P25755 PID:45710 percent identity:

57.11; identified by sequence similarity; putative"

/codon\_start=1

/transl\_table=11  
/product="thiophene and furan oxidation protein"  
/protein\_id="AAFI0590.1"  
/db\_xref="GI:6458744"

/translation="MTFARSSRSGLOTTAAVATAPGSAGVGVVRVSGPRALSIADGL  
FRKRRSATPGGRFLFGHDADELLDEGLCLVFRGPRSYTGEDVAVRTHGSPAVL  
SRVLAITDLGARLARGFTTRAYLAIGLIDLTQAEAVLGEVASTETARRQALGLS  
GALSERVAGVARALTTLAAIQALLDYPEEGVDEDRAPLLGAQSALELLASAHAG  
RSTRCARLALIGRNAGKSSLNALLGVRSIVTPLGTTTRDYLEAGLELAGVPVTL  
VTAGLURETDLIEAAGVQAVALAGNADLVLVLEDSQPREPLPELPTATVRLR  
TKADLPAAWTDPAALDVSAVTGAGLPAURAEIGAALLGDATGEAMLTPTRQDITVRR  
ALTHVQAARTLPDDLAGELEELHALAEILTQGDQVEDVVDVAFVFRFCVGK"

6131..6928

/gene="DR1017"

6131..6928

/gene="DR1017"

/notes="similar to PID:1653441 percent identity: 51.76;

identified by sequence similarity; putative"

/codon\_start=1

/transl\_table=11

/product="rRNA methylase"

/protein\_id="AAFI0591.1"

/db\_xref="GI:6458745"

/translation="MPDSLPHVITSQNQVQKRLVRLTRRRERDGMIEGARELM  
RAESAGVTVPPELFCPELSPPEARLEFPVGTATLSRAAFKVSRENPDGVLGVAP  
LPAPRLPQGDITVVVVLHGLEKPGNVGAILRTADAAGAGVLVLRGADPGYGNVIR  
ASQAVFSVPTAVLGEDAALGWLAEHNFRTVACTPDAPADYWDTPLTGRVALLGTGH  
AGLPAEWRAAHQAQVRVPMGQADSLNVATAALVLYEALRQRRPADA"

7131..8327

/gene="DR1018"

7131..8327

/gene="DR1018"

/notes="identified by Glimmer2; putative"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAFI0596.1"

/db\_xref="GI:6458750"

/translation="MSAPSPASQDPYRDLRTRLGREFCVREADOLIOAAVORRGWQ  
AVRVLGPRDVAVLODYQMRDENGENDRADWLOASVDLNLAEITVPVPTPEDV  
LURPAPRPVANGRRADLPLMLARVHEIATRSIRASIRTDADLAPLERLRAAEWDVQ  
ATQAEVRWETEDMLAQLSQHARTVTDQVASARQVLELTLVRELAALTGERSSP  
AKLAHNLMTQTRSFLEADPAFLVNEQADQPLTLDIDLRHARYSLGVPLHPNVLRA  
RYGLNYALWQAGELGERAVFVQAEALAAQAEAAQLRTTIEAARTHOSITKQLAA  
HYELEERALKLSRVGDDLSLARVRLEWRQARAAARIQANRLEEAVALLEALVSDSA  
EFTSLG"

8492..10060

/gene="DR1019"

8492..10060

/gene="DR1019"

/notes="similar to GB:AE000783 percent identity: 68.44;

identified by sequence similarity; putative"

/codon\_start=1

/transl\_table=11

/product="glycerol-3-phosphate dehydrogenase"

/protein\_id="AAFI0592.1"

/db\_xref="GI:6458746"

/translation="MTQSGSVDRPAQVRAALAPQEWVDIVVGGSGSLGSAVEAATR  
GHSVLLLEGHDYAKGTSSTKLVHGVRYLAQGNVSLVREALHERGLLRKNAPHLVR  
DKGVFPYQDWAAPFYGILKMYDLAGKLNIGSSYLDRLDKVLRSVSTYKKGGLRG  
GILYFDQGDARLAVTLRLTLEDGFGVALNAPVIGLIKIDGQVGRVFRFDSGQE  
HEARAKIVVNATGVWDDITRMEPTVKPMSPSQGVHVVEKKFPLPGDSAIMTPTD  
DGRVLPAVNHQHVHVI GTTTPVQADFEFHLEEEVEFILTARTGYMDPAGRADVL  
SVYAGRLPVNKEKSGAGSTAALSRDHVRISKGNLITLTGGKWTYRMWGSDTLNR  
AEKLGGLPRLTITPGLHLHGWSEDRPDHWKVIGNDAERVKQPGADQHPDLFPYS  
EAEKLWAVRHESATVEDVLARTSRILLIGARASMEAPRAAAILAEELGKDAAWQAO  
QVEAYRQVAGSYVL"

complement(10120..12087)

/gene="DR1020"

complement(10120..12087)

/gene="DR1020"

/notes="similar to PID:1652085 percent identity: 67.48;

identified by sequence similarity; putative"







```

LOCUS      BX842584                244800 bp      DNA      circular BCT 21-NOV-2003
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 13/13.
ACCESSION  BX842584 AL021426 AL022076 AL022120 AL022121 AL123456 Z80343 Z83864
VERSION    Z94121 Z97188
KEYWORDS   BX842584.1 GI:38490370
SOURCE     complete genome.
ORGANISM   Mycobacterium tuberculosis H37Rv
            Mycobacterium tuberculosis H37Rv
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
            tuberculosis complex.
REFERENCE   1
AUTHORS    Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
            Harris,K., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
            Tekala,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
            Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,
            Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
            Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
            Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
            Squares,S., Squires,R., Sulston,J.E., Taylor,K., Whitehead,S. and
            Barrall,B.G.
TITLE       Deciphering the biology of Mycobacterium tuberculosis from the
            complete genome sequence
JOURNAL    Nature 393 (6685), 537-544 (1998)
MEDLINE    98295987
PUBMED     9634230
REFERENCE   2
AUTHORS    Camus,J.C., Pryor,M.J.J., Medigue,C. and Cole,S.T.
TITLE       Re-annotation of the genome sequence of Mycobacterium tuberculosis
            H37Rv
JOURNAL    Microbiology (Reading, Engl.) 148 (Pt 10), 2967-2973 (2002)
MEDLINE    22255591
PUBMED     12368430
REFERENCE   3 (bases 1 to 244800)
AUTHORS    Parkhill,J.
TITLE       Direct Submission
            Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
            tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
            Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
            Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
            75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
            On or before Nov 21, 2003 this sequence version replaced
            gi:3261511, gi:3256026, gi:3261558, gi:3261559, gi:3261648,
            gi:3261687, gi:3261736, gi:3261805.
            Notes:
            Details of M. tuberculosis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/).
FEATURES   source
            1. .244800
               /organism="Mycobacterium tuberculosis H37Rv"
               /mol_type="genomic DNA"
               /strain="H37Rv"
               /db_xref="taxon:83332"
               complement(89..1396)
               /locus_tag="Rv3722c"
               complement(89..1396)
               /locus_tag="Rv3722c"
               /functions="UNKNOWN"
               /note="Rv3722c, (MTV025.070c), len: 435 aa. Conserved
               hypothetical protein, equivalent to O69513|MLCB2407.14
               (alias Q9CB73|ML2336, 463 aa) HYPOTHETICAL 46.8 KDA
               PROTEIN from Mycobacterium leprae (426 aa), FASTA scores:
               opt: 2505, E(): 8.3e-154, (87.25% identity in 424 aa
               overlap). Also highly similar to Q9RU17|DR1579 CONSERVED
               HYPOTHETICAL PROTEIN from Deinococcus radiodurans (452
               aa), FASTA scores: opt: 1162, E(): 3.1e-67, (44.8%
               identity in 422 aa overlap); and partially similar to
               Q91371|PA1654 PROBABLE AMINOTRANSFERASE from Pseudomonas
               aeruginosa (388 aa) FASTA scores: opt: 162, E(): 0.0078,
               (25.85% identity in 348 aa overlap) and other
               aminotransferases. Tbpase score is 0.900. N-terminus
               extended since first submission (previously 408 aa)."

```

```

/codon_start=1
/evidence=experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAAL8044.2"
/db_xref="GI:38490371"
/translations="MSFDSLSPQELAAHARHOODYAALQCMKALDLTRGKPSAEQL
DUSNQLLSLPDDYRPEGTDTNRYYGGHGLPGLRAIFALLGLIAPNLIAKNNSSLE
LHMDIVFLMYGVDSPPRMIQEQGIKFLCPVPGVDRHFAITETGIMIPIMLQ
DGDVDLIELVADPAIKGMTPVPVFGNPSGVTVSWETVRLVQMTAAADPDLFLPD
NAYAVHTLIDFPROVDVLGLAAKAGNPNRPYPVFASTSKITFAGGVSFFGGSLGNIA
WYLOYAGKKSIGCDKNQLRHLFFGDAGDGVRLHMLRHQOILAPKFKALVAEVLDRLS
ESKIASWTEPKGTFISLDVPGTARTVALAKDVGIAVTEAGASFPYRKDPDDKNIR
IAPSPFSPDNRNAVGLATCALLAATETLLNQLGLASSAPNVR"
1613..1698
/gene="trnA-Ser (GGA)"
1613..1698
/gene="trnA-Ser (GGA)"
/product="trnA-Ser"
/notes="codon recognized: UCC"
/anticodon=(pos:1647..1649,aa:Ser)
1804..2568
/locus_tag="Rv3723"
1804..2568
/locus_tag="Rv3723"
/function="UNKNOWN"
/notes="Rv3723, (MTV025.071), len: 254 aa. Probable
conserved transmembrane protein, with hydrophobic
stretches at the N-terminus, and equivalent to
O69512|ML2337|MLCB2407.13c PUTATIVE MEMBRANE PROTEIN from
Mycobacterium leprae (250 aa), FASTA scores: opt: 1029,
E(): 1.2e-44, (64.45% identity in 253 aa overlap). Tbpase
score is 0.900."
/codon_start=1
/transl_table=11
/product="PROBABLE CONSERVED TRANSMEMBRANE PROTEIN"
/protein_id="CAAL8045.1"
/db_xref="GI:2960147"
/translations="MGRKAVLWHSFSGIGVLYFYVFLPRWPELMGDTGSHLGTGL
WMLSDAAGQWLFGIYCAAAAIAGLFGFGLVFAELPPPPKLPKPKPKORLRR
KTKAKGDEAPEAEAEANTELAQDEEAVEAPPESIESPGGEPESATREAPAEETA
TAEPRGGLNRRPTGKTSRRRTRRSGVQVAKVDE"
2735..2977
/gene="cut5a"
/locus_tag="Rv3724a"
2735..2977
/locus_tag="cut5a"
/gene="cut5a"
/EC_number="3.1.1.-"
/function="HYDROLYSIS OF CUTIN (A POLYESTER THAT FORMS THE
STRUCTURE OF PLANT CUTICLE)"
/notes="Rv3724a, (MTV025.072), len: 80 aa. Probable cut5a,
truncated cutinase precursor (EC 3.1.1.-), similar to
N-terminal end of others e.g. Q9KK87 SERINE ESTERASE
CUTINASE from Mycobacterium avium (220 aa), FASTA scores:
opt: 202, E(): 1.5e-06, (56.45% identity in 62 aa
overlap); Q9XB09|RVD2-RV1758 PROTEIN (FRAGMENT) from
Mycobacterium bovis BCG (143 aa), FASTA scores: opt: 200,
E(): 1.5e-06, (61.4% identity in 57 aa overlap); and
Q00298|CUTI_BOTCI|CUTA CUTINASE PRECURSOR from Botrytis
cinerea (Botryotinia fuckeliana) (202 aa), FASTA scores:
opt: 108, E(): 2.2, (40.4% identity in 52 aa overlap).
Also highly similar to others from Mycobacterium
tuberculosis e.g.
O06318|CUT3_MYCTU|Rv3451|MT3557|MTCY13E12.04 PROBABLE
CUTINASE PRECURSOR (247 aa), FASTA scores: opt: 189, E():
1.2e-05, (58.0% identity in 50 aa overlap);
Q50664|CUT2_MYCTU|Rv2301|MT2358|MTCY339.08c PROBABLE
CUTINASE PRECURSOR (219 aa), FASTA scores: opt: 172, E():
0.00015, (59.2% identity in 49 aa overlap);
O06793|Rv1758|MTCY28.24|Z95890 HYPOTHETICAL 17.9 KDA
PROTEIN (174 aa), FASTA scores: opt: 641, E(): 2.7e-29,

```

(57.2% identity in 166 aa overlap);  
 O06319|RV3452|MTV13E12.05; and U00015.11 from *Mycobacterium leprae*. BELONGS TO THE CUTINASE FAMILY. Rest of cutinase ORF continues as RV3724B|CUT5B, frameshifting could occur near position 4169668. Sequence has been checked but no errors found."

/codon\_start=1  
 /transl\_table=11  
 /product="PROBABLE CUTINASE PRECURSOR [FIRST PART] CUT5A"  
 /protein\_id="CAE55632.1"  
 /db\_xref="GI:38490372"  
 /translation="MDVIRWARLLAVVAGTAAAVTTTGLLSAHVPMVMSABPCPDVEVV  
 FARGTGPPPGIGSVGGFLVDALRFPQWRQVTRGLRR"

gene

2874..3437

/gene="cut5b"

/locus\_tag="RV3724B"

2874..3437

/gene="cut5b"

/locus\_tag="RV3724B"

/EC\_number="3.1.1.-"

/function="HYDROLYSIS OF CUTIN (A POLYESTER THAT FORMS THE

STRUCTURE OF PLANT CUTICLE)."

/notes="RV3724B, (MTV025.072), len: 187 aa. Probable cut5b,

truncated cutinase (EC 3.1.1.-), similar to C-terminal end

of others e.g. O9XB09|RVD2-RV1758 PROTEIN (FRAGMENT) from

*Mycobacterium bovis* BCG (143 aa) FASTA scores: opt: 335,

E(): 3.4e-12, (53.25% identity in 92 aa overlap); O9K87

SERINE ESTERASE CUTINASE from *Mycobacterium avium* (220

aa), FASTA scores: opt: 251, E(): 2.5e-07, (44.05%

identity in 168 aa overlap). Also similar to proteins from

*Mycobacterium tuberculosis* e.g. O06793|RV1758|MTCY28.24

HYPOTHETICAL 17.9 KDA PROTEIN (174 aa) FASTA scores: opt:

641, E(): 2.5e-29, (57.25% identity in 166 aa overlap);

O06319|RV3452|MTV13E12.05 HYPOTHETICAL 23.1 KDA PROTEIN

(226 aa), FASTA scores: opt: 385, E(): 7.5e-15, (46.65%

identity in 165 aa overlap);

O06318|CUT3\_MVCTU|RV3451|MT3557|MTCY13E12.04 PROBABLE

CUTINASE PRECURSOR (247 aa), FASTA scores: opt: 307, E():

1.9e-10, (40.7% identity in 167 aa overlap);

Q10837|CUT1\_MVCTU|RV1984c|MT2037|MTCY39.35 PROBABLE

CUTINASE PRECURSOR (217 aa), FASTA scores: opt: 261, E():

6.7e-08, (50.9% identity in 169 aa overlap); etc; and

U00015.11 from *Mycobacterium lepra*. 5'-end of gene is

RV3724B|CUT5A; frameshifting may occur near position

4169668. TParse score is 0.918."

/codon\_start=1

/transl\_table=11

/product="PROBABLE CUTINASE [SECOND PART] CUT5B"

/protein\_id="CAE55633.1"

/db\_xref="GI:38490373"

/translation="MAPGSHLVLAASEDCSSHCVSQVGAISLVGVAVNYPASNDPAS

SDPFKTVIDIRAGSHQISWAMSCPTQRLVGGYSQGAAGVTVSAVVPVAVQQA

VPAPMAPEVAVHVAATLVFGAPSAFLGQYGAIPATIGPLYPKTLQLCADGSDSICGD

GNSPVAHGLYAVNGVWGQGANFAASRL"

## Alignment Scores:

Pred. No.:	165	Length:	244800
Score:	108.00	Matches:	46
Percent Similarity:	35.42%	Conservative:	22
Best Local Similarity:	23.96%	Mismatches:	52
Query Match:	10.39%	Indels:	72
DB:	1	Gaps:	8

US-10-691-383-2\_COPY\_435\_632 (1-198) x BX842584 (1-244800)

Qy	4	GlyThrSerHisTyr-----PheArgLeuIleGlyAla 14
Db	103548	GGACTCTCCACATTCGGCGAGCACTGCATCGTTGGCTGATCCTGGCGCTGCTGGCGCG 103489
Qy	15	AlaGluLeuAlaGlnAlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArg 34
Db	103488	ATAGCCCTGCCACCGCGCCG-----CGGAATGCTTGTG----- 103453

Qy	35	ProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPhe 54
Db	103452	---GCGGGCGCGCGCGTTCGTGCGCCCATCGCC----- 103417
Qy	55	AspIleSerLeuLeuGluAsnAspGluLeuLeuLeuArgValAlaGluIleAsnAlaAla 74
Db	103416	-----GTGCTGATCAAGCGCTGCTGCGG-----CGT 103390
Qy	75	GlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySerPro 94
Db	103389	CAGCGCGCGGATCATCCGGCCATC-----GCGGTCAACGTGGACACGCCA 103345
Qy	95	ThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeu 114
Db	103344	AGTCAACTGAGCTTTCGTGCGCACACGCCACCTCGACACCGCGCGGCTCTCATG 103285
Qy	115	LysAlaLeuIleGlyLeuAspArgGlyGluCysPheProAsnProValPheProSer 134
Db	103284	GGCAGAGCCACCGGGCTG-----CCGCTACCGGTTGTG----- 103252
Qy	135	AspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIle 154
Db	103252	----- 103252
Qy	155	AsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPhe 174
Db	103251	-----CTGTGCGCGCGCGATGGCGCTGTCGCGAATACTGTGGGGGTCCACTACCCCACT 103198
Qy	175	AspGlyIleGlnGlyLeuLeuGlyGluThrIle 186
Db	103197	GATGTGGCGCGTGGGTGTGTTGCTCTGGCGCGCACCGTC 103162

Search completed: September 19, 2004, 07:32:13

Job time : 2540.17 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2004, 03:52:28 ; Search time 8.6087 Seconds  
(without alignments)  
1187.397 Million cell updates/sec

Title: US-10-691-383-2\_COPY\_435\_632

Perfect score: 1039

Sequence: 1 VNFGTSHYFRLIGAELAQR.....GLLGGTTTRTLHQLMTF 198

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PTUS COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1039	100.0	676	3	US-09-151-189-2
2	1039	100.0	676	4	US-09-596-794-2
3	122	11.7	22	3	US-09-151-189-9
4	122	11.7	22	4	US-09-596-794-9
5	110	10.6	21	4	US-09-596-794-11
6	108	10.4	19	3	US-09-151-189-10
7	108	10.4	19	4	US-09-596-794-10
8	99	9.5	19	3	US-09-151-189-11
9	87.5	8.4	605	4	US-09-832-496-2
10	87.5	8.4	605	4	US-09-832-616-2
11	87	8.4	604	4	US-09-832-615-2
12	87	8.4	604	4	US-09-832-617-2
13	85.5	8.2	609	3	US-09-271-778-1
14	85.5	8.2	609	4	US-09-788-871-1
15	85.5	8.2	615	4	US-09-252-991A-28135
16	83.5	8.0	600	2	US-08-679-405-2
17	83.5	8.0	600	2	US-08-842-799-2
18	83.5	8.0	600	3	US-09-271-778-2
19	83.5	8.0	600	4	US-09-788-871-2
20	83.5	8.0	600	5	FCT-US96-11458-2
21	82.5	7.9	290	4	US-09-134-001C-3273
22	81.5	7.8	599	4	US-09-832-441-2
23	81.5	7.8	599	4	US-09-833-102-2
24	80.5	7.7	386	4	US-09-732-615-4
25	78.5	7.6	480	4	US-09-252-991A-32628
26	78.5	7.6	697	4	US-09-489-039A-7485
27	75.5	7.3	244	3	US-08-750-145A-24

Sequence 28, Appl  
Sequence 28, Appl  
Sequence 28, Appl  
Sequence 246, App  
Sequence 27619, A  
Sequence 13, Appl  
Sequence 33097, A  
Sequence 2, Appl  
Sequence 12, Appl  
Sequence 30391, A  
Sequence 17417, A  
Sequence 6983, A  
Sequence 18842, A  
Sequence 4962, Ap  
Sequence 5493, Ap  
Sequence 28726, A  
Sequence 5912, Ap

#### ALIGNMENTS

##### RESULT 1

US-09-151-189-2  
; Sequence 2, Application US/09151189  
; Patent No. 6232457  
; GENERAL INFORMATION:  
; APPLICANT: Vreeland, Valerie  
; APPLICANT: Ng, Kwan L.  
; APPLICANT: The Regents of the University of California  
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses  
; FILE REFERENCE: 023070-087100US  
; CURRENT APPLICATION NUMBER: US/09/151.189  
; CURRENT FILING DATE: 1998-09-10  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 676  
; TYPE: PRT  
; ORGANISM: Fucus distichus  
US-09-151-189-2

Query Match 100.0%; Score 1039; DB 3; Length 676;  
Best Local Similarity 100.0%; Pred. No. 1.6e-119;  
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	VNFGTSHYFRLIGAELAQRASCYQKQVHRFARPEALGGTLHNTIAGDLDDADFDISLLE	60
Db	435	VNFGTSHYFRLIGAELAQRASCYQKQVHRFARPEALGGTLHNTIAGDLDDADFDISLLE	494
Qy	61	NDELLKRVAEINAQNPNNEVTYLLPOAIOVGSPTSPSGHATQNGAFATVLKALIGL	120
Db	495	NDELLKRVAEINAQNPNNEVTYLLPOAIOVGSPTSPSGHATQNGAFATVLKALIGL	554
Qy	121	DRGCEFPNPFVDDGLELINFEGACLTVEGINKLVANVAFQRMGLIHRYDFDGIQGL	180
Db	555	DRGCEFPNPFVDDGLELINFEGACLTVEGINKLVANVAFQRMGLIHRYDFDGIQGL	614
Qy	181	LLGTTITRTLHQLMTTF 198	
Db	615	LLGTTITRTLHQLMTTF 632	

##### RESULT 2

US-09-596-794-2  
; Sequence 2, Application US/09596794  
; Patent No. 6656715  
; GENERAL INFORMATION:  
; APPLICANT: Vreeland, Valerie  
; APPLICANT: The Regents of the University of California  
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases and Their Uses

```
; FILE REFERENCE: 023070-087110US
; CURRENT APPLICATION NUMBER: US/09/596,794
; CURRENT FILING DATE: 2000-06-19
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 676
; TYPE: PRT
; ORGANISM: Fucus distichus
US-09-596-794-2

Query Match      100.0%; Score 1039; DB 4; Length 676;
Best Local Similarity 100.0%; Pred. No. 1.6e-119;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNFGTSHYFRLGAAELAQASCYQKWQVHRFARPEALGCTLHNTIAGDLDAFDISLLE 60
DB 435 VNFGTSHYFRLGAAELAQASCYQKWQVHRFARPEALGCTLHNTIAGDLDAFDISLLE 494
QY 61 NDELLKRVASINAAQNPNNEVTYLLPOAIOVGSPTSPSGHATONGAPATVVKALIGL 120
DB 495 NDELLKRVASINAAQNPNNEVTYLLPOAIOVGSPTSPSGHATONGAPATVVKALIGL 554
QY 121 DRGCECFPNVFPDGLLEINFEAGCLTYEGEINKLVANVAFGRQMLGIHYRFDGIQGL 180
DB 555 DRGCECFPNVFPDGLLEINFEAGCLTYEGEINKLVANVAFGRQMLGIHYRFDGIQGL 614
QY 181 LLGETTIVRTLHQLMTF 198
DB 615 LLGETTIVRTLHQLMTF 532

RESULT 3
US-09-151-189-9
; Sequence 9, Application US/09151189
; Patent No. 6232457
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: Ng, Kwan L.
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses
; FILE REFERENCE: 023070-087100US
; CURRENT APPLICATION NUMBER: US/09/151,189
; CURRENT FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: conserved
; OTHER INFORMATION: vanadium-binding region 1, amino acids 452-473
US-09-151-189-9

Query Match      11.7%; Score 122; DB 3; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.7e-08;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 AQRASYQKWQVHRFARPEALG 39
DB 1 AQRASYQKWQVHRFARPEALG 22

RESULT 4
US-09-596-794-9
; Sequence 9, Application US/09596794
; Patent No. 6656715
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases
; FILE REFERENCE: 023070-087100US
; CURRENT APPLICATION NUMBER: US/09/596,794
; CURRENT FILING DATE: 2000-06-19
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: conserved Fucus
; OTHER INFORMATION: vanadium-binding region 3, amino acids 591-609,
; OTHER INFORMATION: 3rd conserved motif
US-09-596-794-9

Query Match      10.6%; Score 110; DB 4; Length 21;
Best Local Similarity 100.0%; Pred. No. 7.8e-07;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 NKLVANVAFGRQMLGIHYRFD 175
DB 1 NKLVANVAFGRQMLGIHYRFD 21

RESULT 6
US-09-151-189-10
; Sequence 10, Application US/09151189
; Patent No. 6232457
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: Ng, Kwan L.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses
; FILE REFERENCE: 023070-087100US
```

; CURRENT APPLICATION NUMBER: US/09/151,189  
; CURRENT FILING DATE: 1998-09-10  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 19

; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: conserved  
; OTHER INFORMATION: vanadium-binding region 2, amino acids 528-546  
US-09-151-189-10

Query Match 10.4%; Score 108; DB 3; Length 19;  
Best Local Similarity 100.0%; Pred. No. 1.2e-06;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 94 PTHPSYPSGHATONGAFAT 112  
Db 1 PTHPSYPSGHATONGAFAT 19

## RESULT 7

US-09-596-794-10  
; Sequence 10, Application US/09596794  
; Patent No. 6656715  
; GENERAL INFORMATION:  
; APPLICANT: Vreeland, Valerie  
; TITLE OF INVENTION: The Regents of the University of California  
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases  
; FILE REFERENCE: 023070-087110US  
; CURRENT APPLICATION NUMBER: US/09/596,794  
; CURRENT FILING DATE: 2000-06-19  
; PRIOR APPLICATION NUMBER: US 09/151,189  
; PRIOR FILING DATE: 1998-09-10  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: conserved Fucus  
; OTHER INFORMATION: vanadium-binding region 2, amino acids 528-546,  
; OTHER INFORMATION: 2nd conserved motif  
US-09-596-794-10

Query Match 10.4%; Score 108; DB 4; Length 19;  
Best Local Similarity 100.0%; Pred. No. 1.2e-06;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 94 PTHPSYPSGHATONGAFAT 112  
Db 1 PTHPSYPSGHATONGAFAT 19

## RESULT 8

US-09-151-189-11  
; Sequence 11, Application US/09151189  
; Patent No. 6232457  
; GENERAL INFORMATION:  
; APPLICANT: Vreeland, Valerie  
; APPLICANT: Ng, Kwan L.  
; TITLE OF INVENTION: The Regents of the University of California  
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses  
; FILE REFERENCE: 023070-087100US  
; CURRENT APPLICATION NUMBER: US/09/151,189  
; CURRENT FILING DATE: 1998-09-10  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 19

; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: conserved  
; OTHER INFORMATION: vanadium-binding region 3, amino acids 591-609  
US-09-151-189-11

Query Match 9.5%; Score 99; DB 3; Length 19;  
Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 157 LAVNVAFGRQMLGIHYRFD 175  
Db 1 LAVNVAFGRQMLGIHYRFD 19

## RESULT 9

US-09-832-496-2  
; Sequence 2, Application US/09832496  
; Patent No. 6503508  
; GENERAL INFORMATION:  
; APPLICANT: Danielssen, Steffen  
; APPLICANT: Schneider, Palle  
; TITLE OF INVENTION: Polypeptides having haloperoxidase activity  
; FILE REFERENCE: 10042.200-US  
; CURRENT APPLICATION NUMBER: US/09/832,496  
; CURRENT FILING DATE: 2001-04-11  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 605  
; TYPE: PRT  
; ORGANISM: Dreschlera hartliebii  
US-09-832-496-2

Query Match 8.4%; Score 87.5; DB 4; Length 605;  
Best Local Similarity 23.2%; Pred. No. 0.1;  
Matches 32; Conservative 17; Mismatches 38; Indels 51; Gaps 5;

Qy 71 INAAQPNNEVYLLPQAIQVGSPTSPSPSGHATONGAF-----110  
Db 376 LGPATNTNDIPF-----KPPFPAYSGHATFGAVFQVRRYNGRVGTWDDPEP 426

Qy 111 -----ATVLKALIGLDRGECFPNVPFSDGGLLELI-----NFGAC-LTYEGEINKL 157  
Db 427 DNIAIDMWVSEELNGLSRDLQRDYDPTAPIEDQPIVTRVVRHFNLSAWELMFENAIISRI 486

Qy 158 AVNVAFGRQMLGIHYRFD 175  
Db 487 -----FLGVHWRPD 495

## RESULT 10

US-09-832-616-2  
; Sequence 2, Application US/09832616  
; Patent No. 6506586  
; GENERAL INFORMATION:  
; APPLICANT: Danielssen, Steffen  
; APPLICANT: Schneider, Palle  
; TITLE OF INVENTION: Nucleic acids encoding polypeptides having haloperoxidase activity  
; FILE REFERENCE: 10175.200-US  
; CURRENT APPLICATION NUMBER: US/09/832,616  
; CURRENT FILING DATE: 2001-04-11  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 605  
; TYPE: PRT  
; ORGANISM: Dreschlera hartliebii  
US-09-832-616-2

Query Match 8.4%; Score 87.5; DB 4; Length 605;  
Best Local Similarity 23.2%; Pred. No. 0.1;

Matches 32; Conservative 17; Mismatches 38; Indels 51; Gaps 5;  
Qy 71 INAAQPNNEVTYLLPQAIQVGSPTHPSPYSGHATQNGAF----- 110  
Db 376 LGAPATNTNDIPF-----KPPFPAYPSGHATFGSAVQFQVRRYNGRVGTWKDEPDNID 426  
Qy 111 -----ATVLKALIGLDRGCECPNPVPFSDGGLLELI-----NFGAC-LTYEGEINKL 157  
Db 427 DNTAIDMWVSEELNGLSRDLRQYDPTAPIDQPGIVRTRVVRHFSAWELMFENAIISRI 486  
Qy 158 AVNVAFGQMLGIHYRFD 175  
Db 487 -----FLGVHWRFD 495

## RESULT 11

US-09-832-615-2  
; Sequence 2, Application US/09832615  
; Patent No. 6509181  
; GENERAL INFORMATION:  
; APPLICANT: Danielsen, Steffen  
; APPLICANT: Schneider, Falle  
; TITLE OF INVENTION: Polypeptides having haloperoxidase activity  
; FILE REFERENCE: 10041.200-US  
; CURRENT APPLICATION NUMBER: US/09/832,615  
; CURRENT FILING DATE: 2001-04-11  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 2  
; LENGTH: 604  
; TYPE: PRT  
; ORGANISM: Dendryphiella salina  
US-09-832-615-2

Query Match 8.4%; Score 87; DB 4; Length 604;  
Best Local Similarity 22.4%; Pred. No. 0.12;  
Matches 43; Conservative 21; Mismatches 52; Indels 76; Gaps 8;

Qy 51 DADFDISLLENDELLKRVAEINAAQ-----NPNNEVTYLLPQA----- 88  
Db 318 EADLATSEVNADFARLLALNVASADAGIFSWKEWFEYRPLSGVRDGRPDHADPF 377  
Qy 89 -IQVGSPT-----HPSYPSGHAT-----QNGAFAT----- 112  
Db 378 WLTLGAPATNTNDIPFKPPFPAYPSGHATFGSAVQFQVRRYNGRVGTWKDEPDNID 437  
Qy 113 --VLKALIGLDRGCECPNPVPFSDGGLLELI-----NFGAC-LTYEGEINKL 163  
Db 438 MMVSEELNGLSRDLRQYDPTAPIDQPGIVRTRVVRHFSAWELMFENAIISRI----- 491  
Qy 164 GQMLGIHYRFD 175  
Db 492 ---FLGVHWRFD 500

## RESULT 12

US-09-832-617-2  
; Sequence 2, Application US/09832617  
; Patent No. 6511835  
; GENERAL INFORMATION:  
; APPLICANT: Danielsen, Steffen  
; APPLICANT: Schneider, Falle  
; TITLE OF INVENTION: Nucleic acids encoding polypeptides having haloperoxidase activity  
; FILE REFERENCE: 10174.200-US  
; CURRENT APPLICATION NUMBER: US/09/832,617  
; CURRENT FILING DATE: 2001-04-11  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 2  
; LENGTH: 604  
; TYPE: PRT  
; ORGANISM: Dendryphiella salina  
US-09-832-617-2

Query Match 8.4%; Score 87; DB 4; Length 604;  
Best Local Similarity 22.4%; Pred. No. 0.12;  
Matches 43; Conservative 21; Mismatches 52; Indels 76; Gaps 8;  
Qy 51 DADFDISLLENDELLKRVAEINAAQ-----NPNNEVTYLLPQA----- 88  
Db 318 EADLATSEVNADFARLLALNVASADAGIFSWKEWFEYRPLSGVRDGRPDHADPF 377  
Qy 89 -IQVGSPT-----HPSYPSGHAT-----QNGAFAT----- 112  
Db 378 WLTLGAPATNTNDIPFKPPFPAYPSGHATFGSAVQFQVRRYNGRVGTWKDEPDNID 437  
Qy 113 --VLKALIGLDRGCECPNPVPFSDGGLLELI-----NFGAC-LTYEGEINKL 163  
Db 438 MMVSEELNGLSRDLRQYDPTAPIDQPGIVRTRVVRHFSAWELMFENAIISRI----- 491  
Qy 164 GQMLGIHYRFD 175  
Db 492 ---FLGVHWRFD 500

## RESULT 13

US-09-271-778-1  
; Sequence 1, Application US/09271778  
; Patent No. 6221821  
; GENERAL INFORMATION:  
; APPLICANT: Svendsen, Allan  
; APPLICANT: Jorgensen, Louise  
; TITLE OF INVENTION: Haloperoxidases With Altered pH Profiles  
; FILE REFERENCE: 5516.200-US  
; CURRENT APPLICATION NUMBER: US/09/271,778  
; CURRENT FILING DATE: 1999-03-18  
; EARLIER APPLICATION NUMBER: PA 1998 00374  
; EARLIER FILING DATE: 1998-03-18  
; EARLIER APPLICATION NUMBER: 60/079,228  
; EARLIER FILING DATE: 1998-03-24  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 609  
; TYPE: PRT  
; ORGANISM: Curvularia inaequalis  
US-09-271-778-1

Query Match 8.2%; Score 85.5; DB 3; Length 609;  
Best Local Similarity 21.7%; Pred. No. 0.18;  
Matches 30; Conservative 22; Mismatches 35; Indels 51; Gaps 5;

Qy 71 INAAQPNNEVTYLLPQAIQVGSPTHPSPYSGHATQNGAFATVLK----- 115  
Db 381 LGAPATNTNDIPF-----KPPFPAYPSGHATFGSAVQFQVRRYNGRVGTWKDEPD 431  
Qy 116 -----ALIGLDRGCECPNPVPFSDGGLLELI-----NFGAC-LTYEGEINKL 157  
Db 432 DNTAIDMWVSEELNGLSRDLRQYDPTAPIDQPGIVRTRVVRHFSAWELMFENAIISRI 491  
Qy 158 AVNVAFGQMLGIHYRFD 175  
Db 492 -----FLGVHWRFD 500

## RESULT 14

US-09-788-871-1  
; Sequence 1, Application US/09788871  
; Patent No. 6372465  
; GENERAL INFORMATION:  
; APPLICANT: Svendsen, Allan  
; APPLICANT: Jorgensen, Louise  
; TITLE OF INVENTION: Haloperoxidases With Altered pH Profiles  
; FILE REFERENCE: 5516.200-US  
; CURRENT APPLICATION NUMBER: US/09/788,871  
; CURRENT FILING DATE: 2001-02-20

```
; PRIOR APPLICATION NUMBER: 09/271,778
; PRIOR FILING DATE: 1999-03-18
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 609
; TYPE: PRT
; ORGANISM: Curvularia inaequalis
US-09-788-871-1

Query Match      8.2%  Score 85.5; DB 4; Length 609;
Best Local Similarity 21.7%  Pred. No. 0.18; 35; Indels 51; Gaps 5;
Matches 30; Conservative 22; Mismatches 22; Indels 51; Gaps 5;

Qy 71 INAAQNPNEVYLLPQAIQVSGSPHESYPESHATQNGAFATVLK-----115
Db 381 LGAPATNTDIPF-----KPPFPAYPSGHATFGAVFQVRRYYNGRVGTWKDDEP 431

Qy 116 -----ALGLDRGECFPNPVPSDDGLELI-----NFGAC-LTYEGEINKL 157
Db 432 DNIAIDMMISEELNGVNRDLRQPYDPTAPIEDQPGIVRTRIVRHFDSAWELMFENAI SRI 491

Qy 158 AVNVAFGROMLGIHYRED 175
Db 492 -----FLGVHWRFD 500

RESULT 15
US-09-252-991A-28135
; Sequence 28135, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28135
; LENGTH: 615
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28135

Query Match      8.2%  Score 85.5; DB 4; Length 615;
Best Local Similarity 22.3%  Pred. No. 0.18;
Matches 42; Conservative 20; Mismatches 65; Indels 61; Gaps 7;

Qy 17 LAQRASCYQKQVHRFARPEA-----LGGTLHTNTIAGDLDDADFDISLLENDELL 65
Db 385 VAQRQADQDQHR-HRLPAPEGEGQGVGEQSHLAGTLHRYVAGG-----427

Qy 66 KRVAEINAAQNPNEVYLLPQAIQVG-----SPTHSPSPSGHATQNGAPA 111
Db 428 ---EQRAAAGEDHQVGMQRQRAEAGPGQVEVHLGPDQLRGDQYAOQPHADHAPYHGHDG 484

Qy 112 TVLKALIGLDRGECF-----PNVFPFSDGDLLEINFEGACLTYEGE-INKLAV 159
Db 485 ELANHLVVVGRFRRCFVHLYSPERNEQPATAVEGGDGP---AGACNLRAGAGVKQLQV 540

Qy 160 NVAFROM 167
Db 541 DMEFGYQM 548
```

Search completed: September 19, 2004, 04:00:33  
Job time : 9.6087 secs

***This Page Blank (uspto)***



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2004, 03:55:28 ; Search time 39.4188 Seconds  
(without alignments)  
1613.049 Million cell updates/sec

Title: US-10-691-383-2\_COPY\_435\_632

Perfect score: 1039

Sequence: 1 VNFGTSHYFRLIGAELAQR.....GALLGETITVRLHQLMTF 198

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1342398 seqs, 321133274 residues

Total number of hits satisfying chosen parameters: 1342398

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1039	100.0	676	9	US-09-840-762A-2
2	1039	100.0	676	16	US-10-691-383-2
3	122	11.7	22	9	US-09-840-762A-9
4	122	11.7	22	16	US-10-691-383-9
5	110	10.6	21	16	US-10-691-383-11
6	108	10.4	19	9	US-09-840-762A-10
7	108	10.4	19	16	US-10-691-383-10
8	108	10.4	165	14	US-10-080-170-629
9	108	10.4	165	16	US-10-080-170-629
10	99	9.5	19	9	US-09-840-762A-11
11	93	9.0	525	14	US-10-156-761-15017
12	89	8.6	303	16	US-10-767-701-44002
13	89	8.6	333	12	US-10-425-114-60720
14	88.5	8.5	647	9	US-09-969-680A-16
15	88.5	8.5	647	10	US-09-969-680A-16

Sequence 4, Appli  
Sequence 76, Appl  
Sequence 1337, Ap  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 77965, A  
Sequence 63648, A  
Sequence 23, Appl  
Sequence 23, Appl  
Sequence 66217, A  
Sequence 14, Appl  
Sequence 325, App  
Sequence 6876, Ap  
Sequence 8288, Ap  
Sequence 52831, A  
Sequence 95, Appl  
Sequence 120, App  
Sequence 686, App  
Sequence 185920,  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 2394, Ap  
Sequence 2036, Ap  
Sequence 131715,  
Sequence 210, App  
Sequence 210, App  
Sequence 210, App  
Sequence 210, App  
Sequence 4, Appli

#### ALIGNMENTS

##### RESULT 1

US-09-840-762A-2  
; Sequence 2, Application US/09840762A  
; Patent No. US20020035245A1  
; GENERAL INFORMATION:  
; APPLICANT: Vreeland, Valerie  
; APPLICANT: Ng, Kwan L.  
; APPLICANT: The Regents of the University of California  
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses  
; FILE REFERENCE: 023070-087100US  
; CURRENT APPLICATION NUMBER: US/09/840, 762A  
; CURRENT FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: 09/151,189  
; PRIOR FILING DATE: 2001-04-23  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 676  
; TYPE: PRT  
; ORGANISM: Fucus distichus  
; US-09-840-762A-2

Query Match 100.0%; Score 1039; DB 9; Length 676;  
Best Local Similarity 100.0%; Pred. No. 1.5e-106;  
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 VNFGTSHYFRLIGAELAQRASCYQKWQVHRFAPEALGGLTHTNTIAGDLDDADFDISLLE 60  
Db 435 VNFGTSHYFRLIGAELAQRASCYQKWQVHRFAPEALGGLTHTNTIAGDLDDADFDISLLE 494  
Qy 61 NDELLKRVAEINAAQNPNEVTYLLPOAIOVGSPHTPSYSGHATONGAFATVLKALIGL 120  
Db 495 NDELLKRVAEINAAQNPNEVTYLLPOAIOVGSPHTPSYSGHATONGAFATVLKALIGL 554  
Qy 121 DRGCECFPNVPFSDGGLLEINFEGACLTVEGEINKLVANVAFGRQMLGHIHYRFDGIGQL 180  
Db 555 DRGCECFPNVPFSDGGLLEINFEGACLTVEGEINKLVANVAFGRQMLGHIHYRFDGIGQL 614

```
QY 181 LLGETITVRTLHQLMTF 198
Db 615 LLGETITVRTLHQLMTF 632

RESULT 2
US-10-691-383-2
; Sequence 2, Application US/10691383
; Publication No. US20040110260A1
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 023070-087110US
; CURRENT APPLICATION NUMBER: US/10/691,383
; CURRENT FILING DATE: 2003-10-21
; PRIOR APPLICATION NUMBER: US/09/596,794
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: US 09/151,189
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 676
; TYPE: PRT
; ORGANISM: Fucus distichus
US-10-691-383-2
Query Match 100.0%; Score 1039; DB 16; Length 676;
Best Local Similarity 100.0%; Pred. No. 1.5e-106;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNFQTSHYFLIGAAELAQASCYQKWQVHRFARPEALGCTLHNTIAGDLADFDISLLE 60
Db 435 VNFQTSHYFLIGAAELAQASCYQKWQVHRFARPEALGCTLHNTIAGDLADFDISLLE 494

QY 61 NDELLKRVAEINAAQNPNNEVTLLPQAIQVGSPTSPSYSGHATQNGAFATVKKALIGL 120
Db 495 NDELLKRVAEINAAQNPNNEVTLLPQAIQVGSPTSPSYSGHATQNGAFATVKKALIGL 554

QY 121 DRGECFPNPFVSDGDLLEINFEAGCLTYEGEINKLAVNVAFGRQMLGIHYRFDGIQGL 180
Db 555 DRGECFPNPFVSDGDLLEINFEAGCLTYEGEINKLAVNVAFGRQMLGIHYRFDGIQGL 614

QY 181 LLGETITVRTLHQLMTF 198
Db 615 LLGETITVRTLHQLMTF 632

RESULT 3
US-09-840-762A-9
; Sequence 9, Application US/09840762A
; Patent No. US20020035245A1
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: Ng, Kwan L.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses
; FILE REFERENCE: 023070-087100US
; CURRENT APPLICATION NUMBER: US/09/840,762A
; CURRENT FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 09/151,189
; PRIOR FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:conserved
; OTHER INFORMATION: vanadium-binding region 1, amino acids 452-473

QY 181 LLGETITVRTLHQLMTF 198
Db 615 LLGETITVRTLHQLMTF 632

RESULT 4
US-10-691-383-9
; Sequence 9, Application US/10691383
; Publication No. US20040110260A1
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 023070-087110US
; CURRENT APPLICATION NUMBER: US/10/691,383
; CURRENT FILING DATE: 2003-10-21
; PRIOR APPLICATION NUMBER: US/09/596,794
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: US 09/151,189
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:conserved Fucus
; OTHER INFORMATION: vanadium-binding region 1, amino acids 452-473,
; OTHER INFORMATION: 1st conserved motif
US-10-691-383-9
Query Match 11.7%; Score 122; DB 16; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.2e-06;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 AQRASCYQKWQVHRFARPEALG 39
Db 1 AQRASCYQKWQVHRFARPEALG 22

RESULT 5
US-10-691-383-11
; Sequence 11, Application US/10691383
; Publication No. US20040110260A1
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 023070-087110US
; CURRENT APPLICATION NUMBER: US/10/691,383
; CURRENT FILING DATE: 2003-10-21
; PRIOR APPLICATION NUMBER: US/09/596,794
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: US 09/151,189
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:conserved Fucus
; OTHER INFORMATION: vanadium-binding region 3, amino acids 591-609,
```

```
; OTHER INFORMATION: 3rd conserved motif
US-10-691-383-11

Query Match      10.6%; Score 110; DB 16; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.5e-05;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 155 NKLAVNVAFGRQMLGIHYRED 175
Db 1 NKLAVNVAFGRQMLGIHYRED 21

RESULT 6
US-09-840-762A-10
; Sequence 10, Application US/09840762A
; Patent No. US20020035245A1
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: Ng, Kwan L.
; TITLE OF INVENTION: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses
; FILE REFERENCE: 023070-087100US
; CURRENT APPLICATION NUMBER: US/09/840,762A
; CURRENT FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 09/151,189
; PRIOR FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: conserved
; OTHER INFORMATION: vanadium-binding region 2, amino acids 528-546
US-09-840-762A-10

Query Match      10.4%; Score 108; DB 9; Length 19;
Best Local Similarity 100.0%; Pred. No. 6.6e-05;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 94 PTHPSYPSGHATQNGAFAT 112
Db 1 PTHPSYPSGHATQNGAFAT 19

RESULT 7
US-10-691-383-10
; Sequence 10, Application US/10691383
; Publication No. US20040110260A1
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 023070-087110US
; CURRENT APPLICATION NUMBER: US/10/691,383
; CURRENT FILING DATE: 2003-10-21
; PRIOR APPLICATION NUMBER: US/09/596,794
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: US 09/151,189
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: conserved Fucus
; OTHER INFORMATION: vanadium-binding region 2, amino acids 528-546,
; OTHER INFORMATION: 2nd conserved motif
US-10-691-383-10

Query Match      10.4%; Score 108; DB 16; Length 19;
Best Local Similarity 100.0%; Pred. No. 6.6e-05;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 94 PTHPSYPSGHATQNGAFAT 112
Db 1 PTHPSYPSGHATQNGAFAT 19

RESULT 8
US-10-080-170-629
; Sequence 629, Application US/10080170
; Publication No. US20030129601A1
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 629
; LENGTH: 165
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-080-170-629

Query Match      10.4%; Score 108; DB 14; Length 165;
Best Local Similarity 24.0%; Pred. No. 0.0014;
Matches 46; Conservative 22; Mismatches 52; Indels 72; Gaps 8;

Qy 4 GTSHY-----FRLLGAELAQASCYQKQVHRFAEALGGTLHNTIAGDLDAF 54
Db 20 GLSHFGHCITGLWLTALLGAIALPRRR---REWL-----AGAGAFVAHAIA----- 63
Qy 55 DISLENDELLKRVAEINNAQNNNEVYLLPQAIQVGSPTSPSYPSGHATQNGAFATVL 114
Db 64 -----VLTKRLVR---RQPDHPAI-----AVNVDTPSQSFPSAHATSTTAAALLM 107
Qy 115 KALIGLDRGCEFPNVPFSDDGLLELNFECACITYEINKLAVNVAFGRQMLGIHYRF 174
Db 108 GRATGL-----PLPV-----LVPPVALSRLLGVHYP 136
Qy 175 DGIQGLLLGETI 186
Db 137 DVAVGVALGATV 148

RESULT 9
US-10-080-170-629
; Sequence 629, Application US/10080170
; Publication No. US20040121322A9
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 629
; LENGTH: 165
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
```



APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E  
APPLICANT: Tabaska, Jack E  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53313)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 60720  
LENGTH: 333  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: LIB3587-278-Cl1\_FLI.pbp  
US-10-425-114-60720

Query Match 8.6%; Score 89; DB 12; Length 333;  
Best Local Similarity 24.3%; Pred. No. 0.5;  
Matches 53; Conservative 23; Mismatches 64; Indels 78; Gaps 11;

Qy 11 LIGAAELAQR-ASCYQKQVHRFA-RPEALGGTTLHNTIAGDLADDFDISLLENDELKRV 68  
Db 51 LTGCASAASRGAGSFDPSRVVQLSWRPRF---LHKGFLLDAECDHIALAKOKLEKSMV 107  
Qy 69 AEINAAQNPNNEV-----TVLLP---QAIVQ-----GSP 94  
Db 108 ADNKGSKSVSEVTRTSSGMFEKKQDEVTVTRIERISAWTFLEPGEAQLHYQNGEK 167  
Qy 95 THPSYPSGHATONGA-----FATVVKALIGLDGRGE-CFPNP-----VFPSSD----- 136  
Db 168 YEPHYDYFDKNNQALGCHRIATVLMYLSNVKSGETIFPNAEGLQPKDDTWSDCARN 227  
Qy 137 -----GLELINF-----EGACLYEGE 153  
Db 228 GYAVKPKVGDAALLFFSLHPDSTTDSLSHGSCPVIQEQ 265

RESULT 14  
US-09-965-529-16  
Sequence 16, Application US/09965529  
Publication No. US20020182671A1  
GENERAL INFORMATION:  
APPLICANT: LAL, Preeti  
APPLICANT: YUE, Henry  
APPLICANT: TANG, Y. Tom  
APPLICANT: BANDMAN, Olga  
APPLICANT: BURFORD, Neil  
APPLICANT: AZIMZAI, Valda  
APPLICANT: BAUGHN, Mariah R.  
APPLICANT: LU, Dyung Aina M.  
APPLICANT: PATTERSON, Chandra  
TITLE OF INVENTION: MEMBRANE ASSOCIATED PROTEINS  
FILE REFERENCE: PP-0731 USA  
CURRENT APPLICATION NUMBER: US/09/965,529  
CURRENT FILING DATE: 2001-09-26  
PRIOR APPLICATION NUMBER: 60/149,641; 60/164,203; PCT/US00/22315  
PRIOR FILING DATE: 1999-08-17; 1999-11-09; 2000-08-14  
NUMBER OF SEQ ID NOS: 74  
SOFTWARE: PERL Program  
SEQ ID NO 16  
LENGTH: 647  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. US20020182671A1 1336728CD1  
US-09-965-529-16

Query Match 8.5%; Score 88.5; DB 9; Length 647;  
Best Local Similarity 28.4%; Pred. No. 1.5;

Matches 38; Conservative 18; Mismatches 39; Indels 39; Gaps 7;  
Qy 15 AELAQASCYQKQVHRFA-RPEALGGTTLHNTIAG---DLADDFDISLENDE-----L 64  
Db 94 ARLSAAAVLY-----LSNPE---GTCEPTRAGLWASHADHLLALLESPPKALTPGLSWL 143  
Qy 65 LKRVAEINAAQNPNNEV---EVTYLLPQAIQVSGTHPSYPSGHATONGAFATVVKALIGLD 121  
Db 144 LORMQARAGQGTPTKACVDIPQLLEAVGAGP-----GSAGGVLAALLDHHV 190  
Qy 122 RGGECF---PNPVF 132  
Db 191 RSGSCFHALFSPQY 204

RESULT 15  
US-09-969-680A-16  
Sequence 16, Application US/09969680A  
Publication No. US20030124649A1  
GENERAL INFORMATION:  
APPLICANT: LAL, Preeti; YUE, Henry  
APPLICANT: TANG, Y. Tom; BANDMAN, Olga  
APPLICANT: BURFORD, Neil; AZIMZAI, Valda  
APPLICANT: BAUGHN, Mariah R.; LU, Dyung Aina M.  
APPLICANT: PATTERSON, Chandra  
TITLE OF INVENTION: MEMBRANE ASSOCIATED PROTEINS  
FILE REFERENCE: PP-0731-1 USA  
CURRENT APPLICATION NUMBER: US/09/969,680A  
CURRENT FILING DATE: 2001-10-02  
PRIOR APPLICATION NUMBER: US00/22315  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/149,641  
PRIOR FILING DATE: 1999-08-17  
PRIOR APPLICATION NUMBER: 60/164,203  
PRIOR FILING DATE: 1999-11-09  
NUMBER OF SEQ ID NOS: 74  
SOFTWARE: PERL Program  
SEQ ID NO 16  
LENGTH: 647  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. US20030124649A1 1336728CD1  
US-09-969-680A-16

Query Match 8.5%; Score 88.5; DB 10; Length 647;  
Best Local Similarity 28.4%; Pred. No. 1.5; Mismatches 39; Indels 39; Gaps 7;  
Matches 38; Conservative 18; Mismatches 39; Indels 39; Gaps 7;  
Qy 15 AELAQASCYQKQVHRFA-RPEALGGTTLHNTIAG---DLADDFDISLENDE-----L 64  
Db 94 ARLSAAAVLY-----LSNPE---GTCEPTRAGLWASHADHLLALLESPPKALTPGLSWL 143  
Qy 65 LKRVAEINAAQNPNNEV---EVTYLLPQAIQVSGTHPSYPSGHATONGAFATVVKALIGLD 121  
Db 144 LORMQARAGQGTPTKACVDIPQLLEAVGAGP-----GSAGGVLAALLDHHV 190  
Qy 122 RGGECF---PNPVF 132  
Db 191 RSGSCFHALFSPQY 204

Search completed: September 19, 2004, 04:03:37  
Job time : 41.4188 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2004, 01:11:56 ; Search time 38.0595 Seconds  
(without alignments)  
1469.920 Million cell updates/sec

Title: US-10-691-383-2\_COPY\_435\_632

Perfect score: 1039

Sequence: 1 VNFQTHYFRLGAEALQRA.....GLLGGTITVRLHQLMTF 198

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq\_29Jan04.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1039	100.0	676	3	AAy87469 Fucus van
2	1039	100.0	676	5	AAE17501 Fucus dis
3	332.5	32.0	598	4	AAE05750 Corallina
4	331.5	31.9	598	4	AAE05756 Corallina
5	329.5	31.7	598	2	AAW80550 Haloperox
6	322.5	31.0	598	4	AAE05755 Corallina
7	322	31.0	597	2	AAW80551 Haloperox
8	322	31.0	598	4	AAE05757 Corallina
9	122	11.7	22	5	AAE17502 Fucus dis
10	110	10.6	21	5	AAE17504 Fucus dis
11	108	10.4	19	5	AAE17503 Fucus dis
12	108	10.4	165	5	ABU05978 M. tuberc
13	100	9.6	37	4	AAE05754 Vanadium
14	88.5	8.5	396	4	ABB70948 Drosophila
15	88.5	8.5	647	4	AAE06574 Human pro
16	88.5	8.5	647	4	AAE06574 Human mem
17	88.5	8.5	647	5	AAU74618 Oestrogen
18	88.5	8.5	647	7	ADE31719 Human 646
19	87.5	8.4	605	5	AAE05754 Vanadium
20	87.5	8.4	605	5	AAE05754 Vanadium
21	87	8.4	198	6	ABU50041 Protein e
22	87	8.4	604	5	AAU09943 Dendryphi
23	87	8.4	604	5	AAU09943 Dendryphi
24	86.5	8.3	441	2	AAE0501 A. thalia
25	86	8.3	192	5	ABU05372 M. tuberc

26	86	8.3	192	6	ABU35724 Protein e
27	85.5	8.2	609	2	AAy27441 C. inaequ
28	85.5	8.2	640	2	AAE82249 Chloroper
29	85.5	8.2	920	5	AAm51152 Mouse Her
30	85.5	8.2	926	5	AAm51153 Mouse Her
31	85.5	8.2	1256	3	AAE21206 Mouse Her
32	85.5	8.2	1256	4	AAE62860 Amino aci
33	85.5	8.2	1256	5	AAm51151 Mouse Her
34	85.5	8.2	1273	5	ABP62888 Human pol
35	85.5	8.2	1437	6	ABP98857 Human str
36	85.5	8.2	1819	7	AAE60309 Human pro
37	84	8.1	168	4	AAE92922 C glutami
38	83.5	8.0	175	6	ABU24907 Protein e
39	83.5	8.0	600	2	AAW12042 Curvulari
40	83.5	8.0	600	2	AAW12042 Curvulari
41	83.5	8.0	622	5	ABP64821 Human pro
42	83.5	8.0	626	5	AAU74619 Oestrogen
43	83.5	8.0	626	5	ABG96322 Human ova
44	83.5	8.0	626	5	ABP70108 Human NOV
45	83.5	8.0	647	4	AAE01677 Human gen

## ALIGNMENTS

### RESULT 1

AAy87469  
ID AAy87469 standard; protein; 676 AA.  
AC AAy87469;  
XX  
XX  
DT 03-JUL-2000 (first entry)  
XX  
DE Fucus vanadium bromoperoxidase.  
XX  
XX Vanadium bromoperoxidase; vanadium haloperoxidase; alga; ODA;  
KW o-dianisidine oxidation; vanadium ion; halogenation; antimicrobial agent;  
XX phenolic adhesive production.  
XX  
OS Fucus sp.  
XX  
FH Key Location/Qualifiers  
FT Region 441..676  
/note= "Conserved C-terminal region"  
FT Region 452..473  
/note= "Conserved vanadium-binding region #1"  
FT Region 528..546  
/note= "Conserved vanadium-binding region #2"  
FT Region 591..609  
/note= "Conserved vanadium-binding region #3"  
WO200015771-A1.  
23-MAR-2000.  
27-AUG-1999; 99WO-US019893.  
10-SEP-1998; 98US-00151189.  
(REGC ) UNIV CALIFORNIA.  
Vreeland V, Ng KL;  
WPI; 2000-271414/23.  
N-PSDB; AAA10305.

Fucus vanadium bromoperoxidase polypeptide and coding sequence, useful for halogenating or oxidizing compounds, e.g. as antibacterial agents.

Claim 5; Page 42-44; 46pp; English.

This sequence represents a vanadium bromoperoxidase from species of Fucus. This enzyme is 73.4 kD in mass and catalyses the oxidation of o-

CC dianisidine (ODA) when complexed with a vanadium ion. The cDNA encoding  
 CC the enzyme was isolated from a Fucus 2-cell embryo cDNA library using  
 CC hybridisation probes (AA10306-AA10307) corresponding to the second and  
 CC third regions that are conserved between Curvularia and Ascoplylum  
 CC vanadium haloperoxidase active sites. Fucus vanadium bromoperoxidase  
 CC cDNAs (one full-length, two 5'-truncated) were then cloned into the pET  
 CC LIC (ligation independent cloning) vector for expression in Escherichia  
 CC coli. The vanadium bromoperoxidase proteins were expressed fused to an N-  
 CC terminal thiodoxin tag which optimises correct folding. The Fucus  
 CC vanadium bromoperoxidase has a specific activity several-fold higher than  
 CC other algal vanadium haloperoxidases for which at least partial sequences  
 CC have been reported. Vanadium haloperoxidases can be used in a method for  
 CC enzymatically halogenating a compound, and in a method for enzymatically  
 CC oxidising a compound. They can be used to halogenate various substances,  
 CC including proteins. They can be used to produce epoxides from alkenes,  
 CC halogenated ketones from alkenes, to produce alpha, gamma-halohydrins  
 CC from cyclopropanes, and to produce dihalogenated products from alkenes  
 CC and alkenes. Vanadium haloperoxidases may also be used to oxidise various  
 CC compounds, making them useful in signal generating systems in place of  
 CC horseradish peroxidase. The enzymes can also be used as enzymatic  
 CC antimicrobial agents and in the production of phenolic adhesives  
 XX  
 SQ Sequence 676 AA;

Query Match 100.0%; Score 1039; DB 3; Length 676;  
 Best Local Similarity 100.0%; Pred. No. 4.6e-114;  
 Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VNFQSHYFRLLGAEALQASQYQKQVHRFARPEALGGLTHTNTAGDLDADFDISLLE 60  
 DB 435 VNFQSHYFRLLGAEALQASQYQKQVHRFARPEALGGLTHTNTAGDLDADFDISLLE 494  
 QY 61 NDELLKRVAEINAAQNPNNEVTYLLPQAIQVGSPTSPSHGATONGAFATVLKALIGL 120  
 DB 495 NDELLKRVAEINAAQNPNNEVTYLLPQAIQVGSPTSPSHGATONGAFATVLKALIGL 554  
 QY 121 DRGECFPNVPFSDGDLINFEAGCLTYEGEINKLVANVAFGRQMLGHIHYFDGQGL 180  
 DB 555 DRGECFPNVPFSDGDLINFEAGCLTYEGEINKLVANVAFGRQMLGHIHYFDGQGL 614  
 QY 181 LLGETITVRLTHOELMTF 198  
 DB 615 LLGETITVRLTHOELMTF 632  
 RESULT 2  
 ID AAE17501  
 XX AAE17501 standard; protein; 676 AA.  
 AC AAE17501;  
 XX  
 DT 07-AUG-2003 (revised)  
 DT 01-MAY-2002 (first entry)  
 XX  
 DE Fucus distiches vanadium haloperoxidase.  
 XX  
 KW Vanadium haloperoxidase; vanadium bromoperoxidase; o-dianisidine; ODA;  
 KW epoxide; antimicrobial agent; phenolic adhesive; halohydrin;  
 KW industrial catalysis; enzyme; catalyst.  
 XX  
 OS Fucus distichus.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 439..461  
 FT /note= "Catalytic frame helices (CFH) alpha-helix 1"  
 FT Region 469..481  
 FT /note= "Catalytic frame helices (CFH) alpha-helix 1a  
 FT (alpha1-alpha2 loop)"  
 FT Region 490..493  
 FT /note= "Catalytic frame helices (CFH) alpha-helix 1b  
 FT (alpha1-alpha2 loop)"  
 FT Region 496..509  
 FT /note= "Catalytic frame helices (CFH) alpha-helix 1c

FT Region (alpha1-alpha2 loop)"  
 FT 536..552  
 FT /note= "Catalytic frame helices (CFH) alpha-helix 2"  
 FT Region 554..557  
 FT /note= "Catalytic frame helices (CFH) alpha-helix 2d  
 FT (alpha2-alpha3 loop)"  
 FT Region 584..602  
 FT /note= "Catalytic frame helices (CFH) alpha-helix 3"  
 FT Region 607..631  
 FT /note= "Catalytic frame helices (CFH) alpha-helix 4"  
 XX  
 XX WO200200838-A2.  
 XX  
 XX 03-JAN-2002.  
 XX  
 XX 07-JUN-2001; 2001WO-US018602.  
 XX  
 XX 19-JUN-2000; 2000US-00596794.  
 XX  
 XX (REGC ) UNIV CALIFORNIA.  
 XX  
 XX Vreeland V;  
 XX  
 XX WPI; 2002-154651/20.  
 XX  
 XX N-PSDB; AAD28155.  
 XX  
 XX Isolated nucleic acid for, e.g. to halogenate various substrates  
 XX including proteins, has polynucleotide sequence encoding vanadium  
 XX haloperoxidase polypeptide with catalytic helical frame.  
 XX  
 XX Claim 15; Page 53-56; 56pp; English.  
 XX  
 CC The invention relates to nucleic acid encoding vanadium haloperoxidase  
 CC particularly vanadium bromoperoxidase polypeptide which consists of a  
 CC catalytic helical frame that complexes a vanadium ion and catalyses the  
 CC oxidation of o-dianisidine (ODA). The invention also provides  
 CC recombinantly produced vanadium haloperoxidases. The polypeptides of the  
 CC invention can be fused to other proteins to allow quantification or  
 CC localisation of the linked protein. They can be used to halogenate  
 CC various substrates including proteins; to produce epoxides from alkenes,  
 CC halogenated ketones from alkenes; to produce alpha, gamma-halohydrins  
 CC from cyclopropanes and to produce dihalogenated products from alkenes and  
 CC alkenes. They can also be used in signal generating systems in place of  
 CC horseradish peroxidase; as a component in assays; as enzymatic  
 CC antimicrobial agents; and in the production of phenolic adhesives. The  
 CC polypeptides can be used in industrial catalysis in a variety of  
 CC contexts, e.g. catalyst for halogenation, oxidation, and epoxidation  
 CC reactions. The present sequence is Fucus distiches vanadium  
 CC haloperoxidase. (Updated on 07-AUG-2003 to correct OS field.)  
 XX  
 XX Sequence 676 AA;  
 QY Query Match 100.0%; Score 1039; DB 5; Length 676;  
 DB Best Local Similarity 100.0%; Pred. No. 4.6e-114;  
 QY Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 1 VNFQSHYFRLLGAEALQASQYQKQVHRFARPEALGGLTHTNTAGDLDADFDISLLE 60  
 DB 435 VNFQSHYFRLLGAEALQASQYQKQVHRFARPEALGGLTHTNTAGDLDADFDISLLE 494  
 QY 61 NDELLKRVAEINAAQNPNNEVTYLLPQAIQVGSPTSPSHGATONGAFATVLKALIGL 120  
 DB 495 NDELLKRVAEINAAQNPNNEVTYLLPQAIQVGSPTSPSHGATONGAFATVLKALIGL 554  
 QY 121 DRGECFPNVPFSDGDLINFEAGCLTYEGEINKLVANVAFGRQMLGHIHYFDGQGL 180  
 DB 555 DRGECFPNVPFSDGDLINFEAGCLTYEGEINKLVANVAFGRQMLGHIHYFDGQGL 614  
 QY 181 LLGETITVRLTHOELMTF 198  
 DB 615 LLGETITVRLTHOELMTF 632



```

RESULT 3
AAE05750
ID AAE05750 standard; protein; 598 AA.
XX AC AAE05750;
XX DT 24-SEP-2001 (first entry)
XX DE Corallina officinalis vanadium bromoperoxidase (cVBPO).
XX KW Marine red algae; vanadium bromoperoxidase; VBPO; haloperoxidase;
XX signal generation; analytical system; anti-microbial agent;
XX industrial application; biological material; avidin; metabolite;
XX lipopolysaccharide; human chorionic gonadotropin; luteinising hormone;
XX vitamin; intact cell; drug; narcotic.
XX OS Corallina officinalis.
XX FH Key Location/Qualifiers
FT Active-site 400
FT /note= "Site at which vanadate cofactor binds"
FT Active-site 408
FT /note= "Site at which vanadate cofactor binds"
FT Misc-difference 474
FT /note= "Ala is present at this location in the sequence
FT shown in sequence listing and figure 2 of the
FT specification"
FT Active-site 547
FT /note= "Site at which vanadate cofactor binds"
FT Active-site 553
FT /note= "Site at which vanadate cofactor binds"
XX WO200153494-A2.
XX PN 26-JUL-2001.
XX PD 24-JAN-2001; 2001WO-US002383.
XX PF 24-JAN-2000; 2000US-0177764P.
XX PR (REGC ) UNIV CALIFORNIA.
XX PA Butler A, Carter JN;
XX PI WPI; 2001-457611/49.
XX DR N-PSDB; AAD11050.
XX DT Novel vanadium haloperoxidase, especially vanadium bromoperoxidase,
XX obtained from Corallina officinalis for use in signal generation in
XX analytical systems.
XX PS Claim 9; Fig 1; 68pp; English.
XX CC The present sequence is Corallina officinalis (marine red algae) vanadium
XX bromoperoxidase (cVBPO). The vanadium haloperoxidase, especially vanadium
XX bromoperoxidase (VBPO) is useful for signal generation in analytical
XX systems, and as anti-microbial agents. The VBPO is also useful for
XX industrial applications. The VBPO is also useful for detecting a wide
XX variety of chemical and biological materials such as amino acids,
XX peptides, polypeptides, proteins including enzymes, avidin, antibodies
XX and antigenic proteins, carbohydrates including monosaccharides,
XX polysaccharides and lipopolysaccharides, hormones such as human chorionic
XX gonadotropin, thyroxine, follicle stimulating hormone, luteinising hormone,
XX hormone, metabolites such as glucose, lactate and pyruvate,
XX oligonucleotides, nucleic acids, vitamins such as B12 and biotin, intact
XX cells from various organisms including microorganisms and drugs such as
XX narcotics, therapeutic and those abused
XX SQ Sequence 598 AA;

Query Match 32.0%; Score 332.5; DB 4; Length 598;
Best Local Similarity 38.5%; Pred. No. 5.7e-30;

Matches 82; Conservative 38; Mismatches 72; Indels 21; Gaps 7;
Qy 1 VNFQTSYHRLI-GAAELAQASCYQKWVHRFARPEALGGTLH-NTIAGD----LDADF 54
Dy 374 VNFQDAHVLSLVTVEVATKAVRYQKFNHRRLRPEATGGLISVKNIAAEKGSVFPEV 433
Qy 55 DISLENDELLKRVAEINAAQN-----PNNEVTVLLPQATQVGSPTTPSPSGHATONGA 109
Dy 434 DLAVEELEDILKAEISNRKQNIADGDPDPSPFLPQAFAGSGFPFPHSYGSHAVVAGA 493
Qy 110 FATVLKALIG----LDRGGCFNPVPSPDDGELINFEAGACTYEGEINKLVNVAFR 165
Dy 494 CVTILKAFPSNCFIDQVFE-----VDKDEKLWKSSFKGT-LTVAGELNKLADNIAGR 547
Qy 166 QMLGIHYRFDGICGLLGETITVTTLHQELMTF 198
Dy 548 NMAGVHYFSDQFESILLGEQVAIGILEEQSLTY 580

RESULT 4
AAE05756
ID AAE05756 standard; protein; 598 AA.
XX AC AAE05756;
XX DT 24-SEP-2001 (first entry)
XX DE Corallina pilulifera vanadium bromoperoxidase 1 (VBPO1).
XX KW Vanadium bromoperoxidase; VBPO; haloperoxidase; signal generation;
XX analytical system; anti-microbial agent; industrial application;
XX biological material; avidin; metabolite; lipopolysaccharide;
XX human chorionic gonadotropin; luteinising hormone; vitamin; intact cell;
XX drug; narcotic.
XX OS Corallina pilulifera.
XX PN WO200153494-A2.
XX PD 26-JUL-2001.
XX PF 24-JAN-2001; 2001WO-US002383.
XX PR 24-JAN-2000; 2000US-0177764P.
XX PA (REGC ) UNIV CALIFORNIA.
XX PI Butler A, Carter JN;
XX DR WPI; 2001-457611/49.
XX DT Novel vanadium haloperoxidase, especially vanadium bromoperoxidase,
XX obtained from Corallina officinalis for use in signal generation in
XX analytical systems.
XX PS Example 1; Fig 2; 68pp; English.
XX CC The present sequence is Corallina pilulifera vanadium bromoperoxidase 1
XX (VBPO1). The vanadium haloperoxidase, especially vanadium bromoperoxidase
XX (VBPO) is useful for signal generation in analytical systems, and as anti-
XX microbial agents. The VBPO is also useful for industrial applications.
XX The VBPO is also useful for detecting a wide variety of chemical and
XX biological materials such as amino acids, peptides, polypeptides,
XX proteins including enzymes, avidin, antibodies and antigenic proteins,
XX carbohydrates including monosaccharides, polysaccharides and
XX lipopolysaccharides, hormones such as human chorionic gonadotropin,
XX thyroid stimulating hormone, luteinising hormone, thyroxine, follicle
XX stimulating hormone, parathyroid hormone and growth hormone, metabolites
XX such as glucose, lactate and pyruvate, oligonucleotides, nucleic acids,
XX vitamins such as B12 and biotin, intact cells from various organisms
XX including microorganisms and drugs such as narcotics, therapeutic and
XX those abused
XX CC

```

```
SQ Sequence 598 AA;
Query Match 31.9%; Score 331.5; DB 4; Length 598;
Best Local Similarity 40.1%; Pred. No. 7.5e-30;
Matches 85; Conservative 34; Mismatches 74; Indels 19; Gaps 8;

Qy 1 VNFGTSHYFRLI-GAAELAQKASCYQKWQVHRFARPEALGGTLH-NTIA-----GDLDAADF 54
Db 374 VNFQSAHVLSLTVETATKALKAVRYQKFNHRRLRPEATGGLISVKNIAQKGSIFPEV 433

Qy 55 DISLENDELLKRVAEINAAQN-----PNNEVTYLLPOAIQVGSPTSPSPSHGATONGA 109
Db 434 DLAVEELGDLLEKAEISNRKQNIADGDDPDPSPFLPMFAEGSPPHPSYSGSHAVVAGA 493

Qy 110 FATVLKALIGLDGEGCFNPNVPFSD---DGLLEINFEAGCLTYEGEINKLVNVAFRQ 166
Db 494 CVTILKAF--FDSGIEI--DQVFEVDKDEKLVKSSFKGT-LTVAGELNKLADNIAIGRN 548

Qy 167 MLGIHYRFDGIQGLLGGTITVTRLHQLMTF 198
Db 549 MAGVHYFSDQFESLLGGEQVAIGILEQSLTY 580

RESULT 5
AAW80550
ID AAW80550 standard; protein; 598 AA.
XX AC AAW80550;
XX DT 09-DEC-1998 (first entry)
XX DE Haloperoxidase enzyme, CP.BP01.
XX KW Haloperoxidase; enzyme; CP.BP01; CP.BP02; vanadium; halide specificity;
XX KM bromine; iodine.
XX OS Corallina pilulifera.
XX PN JPI0248581-A.
XX PD 22-SEP-1998.
XX PF 06-MAR-1997; 97JP-00070539.
XX PR 06-MAR-1997; 97JP-00070539.
XX PA (SAKA ) OTSUKA PHARM CO LTD.
XX PI WPI; 1998-560733/48.
XX DR N-PSDB; AAV56020.
XX PT New haloperoxidase gene - and corresponding vector, transformed host cell
XX PT and method of preparation.
XX PS Claim 1; Page 10-11; 20pp; Japanese.
XX CC This represents a haloperoxidase enzyme, CP.BP01. The haloperoxidase
XX CC genes encoding the enzymes CP.BP01 and CP.BP02 are isolated from
XX CC Corallina pilulifera. A host cell transformed with a vector containing
XX CC the haloperoxidase genes can be used for the recombinant production of
XX CC the enzyme. The haloperoxidases of the invention are vanadium-dependent
XX CC and have halide specificity of bromine and iodine.
XX CC

SQ Sequence 598 AA;
Query Match 31.7%; Score 329.5; DB 2; Length 598;
Best Local Similarity 40.1%; Pred. No. 1.3e-29;
Matches 85; Conservative 34; Mismatches 74; Indels 19; Gaps 8;

Qy 1 VNFGTSHYFRLI-GAAELAQKASCYQKWQVHRFARPEALGG-TLHNNTIA-----GDLDAADF 54
Db 374 VNFQSAHVLSLTVETATKALKAVRYQKFNHRRLRPEATGGLIIVNKAIPQKGSIFPEV 433

Qy 55 DISLENDELLKRVAEINAAQN-----PNNEVTYLLPOAIQVGSPTSPSPSHGATONGA 109
Db 434 DLAVEELGDLLEKAEISNRKQNIADGDDPDPSPFLPMFAEGSPPHPSYSGSHAVVAGA 493

Qy 110 FATVLKALIGLDGEGCFNPNVPFSD---DGLLEINFEAGCLTYEGEINKLVNVAFRQ 166
Db 494 CVTILKAF--FDSGIEI--DQVFEVDKDEKLVKSSFKGT-LTVAGELNKLADNIAIGRN 548

Qy 167 MLGIHYRFDGIQGLLGGTITVTRLHQLMTF 198
Db 549 MAGVHYFSDQFESLLGGEQVAIGILEQSLTY 580

RESULT 6
AAE05755
ID AAE05755 standard; protein; 598 AA.
XX AC AAE05755;
XX DT 24-SEP-2001 (first entry)
XX DE Corallina officinalis vanadium bromoperoxidase mutant (H480A).
XX KW Marine red algae; vanadium bromoperoxidase; VBPO; haloperoxidase;
XX KM signal generation; analytical system; anti-microbial agent;
XX KW industrial application; biological material; avidin; metabolite;
XX KW lipopolysaccharide; human chorionic gonadotropin; luteinising hormone;
XX KW vitamin; intact cell; drug; narcotic; mutant; variant.
XX OS Corallina officinalis.
XX PN WO200153494-A2.
XX PD 26-JUL-2001.
XX PF 24-JAN-2001; 2001WO-US0002383.
XX PR 24-JAN-2000; 2000US-0177764P.
XX PA (REGC ) UNIV CALIFORNIA.
XX PI Butler A, Carter JN;
XX PT WPI; 2001-457611/49.
XX PT Novel vanadium haloperoxidase, especially vanadium bromoperoxidase,
XX PT obtained from Corallina officinalis for use in signal generation in
XX PT analytical systems.
XX PS Example 5; Page; 68pp; English.
XX CC The present sequence is Corallina officinalis (marine red algae) vanadium
XX CC bromoperoxidase (cVBPO) mutant (H480A). The vanadium haloperoxidase,
XX CC especially vanadium bromoperoxidase (VBPO) is useful for signal
XX CC generation in analytical systems, and as anti-microbial agents. The VBPO
XX CC is also useful for industrial applications. The VBPO is also useful for
XX CC detecting a wide variety of chemical and biological materials such as
XX CC amino acids, peptides, polypeptides, proteins including enzymes, avidin,
XX CC antibodies and antigenic proteins, carbohydrates including
XX CC monosaccharides, polysaccharides and lipopolysaccharides, hormones such
XX CC as human chorionic gonadotropin, thyroid stimulating hormone, luteinising
XX CC hormone, thyroxine, follicle stimulating hormone, parathyroid hormone and
XX CC growth hormone, metabolites such as glucose, lactate and pyruvate,
XX CC oligonucleotides, nucleic acids, vitamins such as B12 and biotin, intact
XX CC cells from various organisms including microorganisms and drugs such as
XX CC narcotics, therapeutic and those abused. Note: The present sequence is
XX CC not shown in the specification but is derived from the wild type
XX CC Corallina officinalis vanadium bromoperoxidase (cVBPO) sequence shown in
```







RESULT 13  
AAE05754  
ID AAE05754 standard; peptide; 37 AA.  
XX  
AC AAE05754;  
XX  
DT 24-SEP-2001 (first entry)  
XX  
DE Vanadium bromoperoxidase (VBPO) consensus sequence #2.  
XX  
KW Marine red algae; vanadium bromoperoxidase; VBPO; haloperoxidase;  
KW signal generation; analytical system; anti-microbial agent;  
KW industrial application; biological material; avidin; metabolite;  
KW lipopolysaccharide; human chorionic gonadotropin; luteinising hormone;  
KW vitamin; intact cell; drug; narcotic.  
XX  
OS Corallina officinalis.  
OS Corallina pilulifera.  
OS Ascophyllum nodosum.  
OS Fucus distichus.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 3 /label= Unknown  
FT Misc-difference 4 /label= Unknown  
FT Misc-difference 7 /label= Unknown  
FT Misc-difference 7 /label= Ile, Leu  
FT Misc-difference 12 /label= Unknown  
FT Misc-difference 14 /label= Unknown  
FT Misc-difference 16 /label= Unknown  
FT Misc-difference 19 /label= Unknown  
FT Misc-difference 21 /label= Unknown  
FT Misc-difference 23 /label= Unknown  
FT Misc-difference 26 /label= Unknown  
FT Misc-difference 27 /label= Unknown  
FT Misc-difference 29 .32 /label= Unknown  
XX  
PN WO200153494-A2.  
XX  
PD 26-JUL-2001.  
XX  
PF 24-JAN-2001; 2001WO-US002383.  
XX  
PR 24-JAN-2000; 2000US-0177764P.  
XX  
PA (REGC ) UNIV CALIFORNIA.  
XX  
XX Butler A, Carter JN;  
XX  
XX WPI; 2001-457611/49.  
XX  
XX Novel vanadium haloperoxidase, especially vanadium bromoperoxidase,  
XX obtained from Corallina officinalis for use in signal generation in  
XX analytical systems.  
XX  
XX Example 1; Page 41; 68pp; English.  
XX  
XX The present sequence is a vanadium bromoperoxidase (VBPO) consensus  
XX sequence which is derived from Corallina officinalis, Corallina  
XX pilulifera, Ascophyllum nodosum and Fucus distichus. The vanadium  
XX haloperoxidase, especially vanadium bromoperoxidase (VBPO) is useful for

CC signal generation in analytical systems, and as anti-microbial agents.  
CC The VBPO is also useful for industrial applications. The VBPO is also  
CC useful for detecting a wide variety of chemical and biological materials  
CC such as amino acids, peptides, polypeptides, proteins including enzymes,  
CC avidin, antibodies and antigenic proteins, carbohydrates including  
CC monosaccharides, polysaccharides and lipopolysaccharides, hormones such  
CC as human chorionic gonadotropin, thyroid stimulating hormone, luteinising  
CC hormone, thyroxine, follicle stimulating hormone, parathyroid hormone and  
CC growth hormone, metabolites such as glucose, lactate and pyruvate.  
CC oligonucleotides, nucleic acids, vitamins such as B12 and biotin, intact  
CC cells from various organisms including microorganisms and drugs such as  
CC narcotics, therapeutic and those abused  
XX  
SQ Sequence 37 AA;  
Query Match 9.8%; Score 100; DB 4; Length 37;  
Best Local Similarity 59.5%; Pred. No. 0.00037;  
Matches 22; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
QY 148 LTYEGEINKLVNVAFGROMLGTHYRFDGIGQLLGE 184  
DB 1 LTYXGXNKLVNVAFGROMLGTHYRFDGIGQLLGE 37  
RESULT 14  
ABB70948  
ID ABB70948 standard; protein; 396 AA.  
XX  
AC ABB70948;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 39636.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR N-PSDB; ABL15051.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
XX Disclosure; SEQ ID NO 39636; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
XX ABB72072). The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ffp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 396 AA;

Query Match 8.5%; Score 88.5; DB 4; Length 396;  
Best Local Similarity 26.5%; Pred. No. 0.35;  
Matches 43; Conservative 20; Mismatches 60; Indels 39; Gaps 6;  
QY 48 GDLADADFDISLENDELLKRVAEINAAQNPNVETLLPOAIQVGSPTHTSPSGHATQ 107  
DB 65 GNLDAALLDAPLNETPANRV-ELYPNKLLNLRSSLLPAGQGDVDPHQPWPNRSITFD 123  
QY 108 GAFATVLKAL-----IGLDRGGCFNPNVFPSPDDGLELINFEGACLTYEGIN 155  
DB 124 GAENVWKTFCQVHCHLRVEAYGADNWGGIYDN---ESSDGM-----LGDY 168  
QY 156 KLVAVAFGRQMLGIHYRFDGIQGLLGLTITVRTLHOELMT 197  
DB 169 EQRVEMAIG---CIYNYWDGI-----TETSHTIARSSVT 199

RESULT 15  
AAE06574  
ID AAE06574 standard; protein; 647 AA.  
XX AAE06574;  
AC AAE06574;  
XX 25-SEP-2001 (first entry)  
XX Human protein having hydrophobic domain, HP10755.

XX Human; hydrophobic domain; gene therapy; nutritional supplement;  
XX cell proliferation; immunomodulatory; autoimmune disorder; antimicrobial;  
XX multiple sclerosis; rheumatoid arthritis; insulin-dependent diabetes;  
XX haematopoiesis; tissue growth activity; Parkinson's disease; cytostatic;  
XX Huntington's disease; Alzheimer's disease; chemotactic; chemokinetic;  
XX haemostatic; thrombolytic; tumour growth inhibitor; anabolic;  
XX contraceptive; antifertility; antiinflammatory.

XX Homo sapiens.  
XX WO200149728-A2.  
XX 12-JUL-2001.  
XX 28-DEC-2000; 2000WO-JP009359.  
XX 06-JAN-2000; 2000JP-00000585.  
XX 08-JAN-2000; 2000JP-00000588.  
XX 11-JAN-2000; 2000JP-00002299.  
XX 03-FEB-2000; 2000JP-00026862.  
XX 03-MAR-2000; 2000JP-00058367.

XX (PROT-) PROTEGENE INC.  
XX (SAGA ) SAGAMI CHEM RES CENT.  
XX Kato S, Kimura T;  
XX WPI; 2001-418355/44.  
XX N-PSDB; AAD12569.

XX Human proteins with hydrophobic domains and the nucleic acids encoding  
XX them, useful for preventing diagnosing and treating e.g. cancer,  
XX Alzheimer's and inflammation.

XX Claim 1; Page 67-68; 563pp; English.  
XX The present sequence is human protein with hydrophobic domain, HP10755.  
XX The polynucleotide and polypeptide of the invention may be used in the  
XX prevention, diagnosis and treatment of diseases associated with  
XX inappropriate polypeptide expression. The polynucleotides may be used to  
XX produce the polypeptide, by inserting the nucleic acids into a host cell  
XX and culturing the cell to express the protein. The polynucleotides and  
XX its complementary sequences may also be used as DNA probes in diagnostic  
XX assays and also used in gene therapy. The polypeptides may also be used  
XX as antigens in the production of antibodies and in assays to identify

CC modulators of polypeptide expression and activity. The polypeptides and  
CC nucleic acids may be used as nutritional supplements, to modulate  
CC cytokine and cell proliferation activity, to modulate immune stimulation  
CC or suppression (e.g. for the treatment of microbial infections and  
CC autoimmune disorders such as multiple sclerosis, rheumatoid arthritis and  
CC insulin-dependent diabetes), to modulate haematopoiesis, to modulate  
CC tissue growth activity (e.g. for the treatment of Parkinson's disease,  
CC Huntington's disease and Alzheimer's disease), to modulate activin and  
CC inhibin activity (e.g. for controlling fertility), to modulate  
CC chemotactic and chemokinetic activity, to modulate haemostatic and  
CC thrombolytic activity, to modulate receptor ligand activity, to modulate  
CC inflammation and to inhibit tumour growth

XX  
SQ Sequence 647 AA;  
Query Match 8.5%; Score 88.5; DB 4; Length 647;  
Best Local Similarity 28.4%; Pred. No. 0.75;  
Matches 38; Conservative 18; Mismatches 39; Indels 39; Gaps 7;  
QY 15 AELAQRASCYQKQVHRFARPEALGGTLHNTIAG--DLADADFDISLENDE-----L 64  
DB 94 ARLSAAAVLY-----LSNPE---GTCTDTRAGLWASHADHLLALLESFKALTPGLSWL 143  
QY 65 LKRVAEINAAQNPN---EVTYLLPOAIQVGSPTHTSPSGHATQNGAFATVLKALIGLD 121  
DB 144 LORMQARAAGQTKTACVDIPQLLEAVGAGAP-----GSAGGVLAALLDHY 190  
QY 122 RGGECEP---PNPVF 132  
DB 191 RSGSCFHALPSPQY 204

Search completed: September 19, 2004, 03:55:20  
Job time : 39.0595 secs

***This Page Blank (uspio)***



Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	116.5	11.2	145	2	F75447	vanadium chloroper
2	108	10.4	165	2	C70888	hypothetical prote
3	87	8.4	198	2	AG0161	probable permease
4	86.5	8.3	441	2	F86185	hypothetical prote
5	86	8.3	192	2	F86920	probable membrane
6	85.5	8.2	609	1	S69334	chloride peroxidase
7	84.5	8.1	217	2	D97672	hypothetical prote
8	84.5	8.1	242	2	AB2897	conserved hypothet
9	84	8.1	418	2	AF2306	hypothetical prote
10	82.5	7.9	330	1	F64346	hypothetical prote
11	82	7.9	567	2	T11653	hypothetical prote
12	80.5	7.7	433	2	AG1890	hypothetical prote
13	80	7.7	489	2	AB2298	mercuric reductase
14	79.5	7.7	198	1	A64822	probable bacitraci
15	79.5	7.7	198	2	AG9744	hypothetical prote
16	79.5	7.7	198	2	E85594	hypothetical prote
17	79.5	7.7	501	2	T34741	probable ABC-type
18	79	7.6	266	2	F70872	hypothetical prote
19	78.5	7.6	228	2	F71886	hypothetical prote
20	78.5	7.6	385	2	T23980	hypothetical prote
21	77.5	7.5	227	2	C64626	conserved hypothet
22	77.5	7.5	454	2	E70448	signal recognition
23	77.5	7.5	940	2	S73950	phosphotransferase
24	77	7.4	260	2	D75404	phosphatidylglycer
25	77	7.4	495	1	MM5CTC	export system oute
26	77	7.4	495	2	C91119	outer membrane cha
27	77	7.4	495	2	B85964	hypothetical prote
28	77	7.4	551	2	D83277	electron transfer
29	76.5	7.4	576	2	A40695	collagen alpha 1(V

A;Cross-references: GB:AL022076; GB:AL123456; NID:G3256026; PIDN:CAA17871.1; PID:el125654  
A;Experimental source: Strain H37Rv  
C;Genetics:  
A;Gene: RV3807c  
C;Superfamily: bacitracin transport permease; glucose-6-phosphatase catalytic domain hom

Query Match 10.4%; Score 108; DB 2; Length 165;  
Best Local Similarity 24.0%; Pred. No. 0.0063;  
Matches 46; Conservative 22; Mismatches 52; Indels 72; Gaps 8;  
Qy 4 GTSHY-----FRILGAELAQASCYQKQVHRPARPEALGGTLHNTIAGDLDAF 54  
Db 20 GLSHFGHCIGWLITALLGALPRR---REWL-----AGAGAFVAHAIA----- 63  
Qy 55 DISLENDELLKRVAEINAAQNPNEVYLLPQAIQVSGTHSPSPGHATQNGAFATVL 114  
Db 64 -----VLTKRLVR---RQPDHPAI-----AVNVDTSPQLSPSAHATSTTAAALLM 107  
Qy 115 KALIGDRGCECPNPVPPSDDGLLEINFEACLYTEGEINKLVNNAFQKMLGIHYRF 174  
Db 108 GRATGL-----PLPV-----LVPPMALSRIILGVHYPS 136  
Qy 175 DGIQGLLGETI 186  
Db 137 DVAUGVALGATV 148

RESULT 3  
AG0161  
probable permease YP01324 [imported] - Yersinia pestis (strain CO92)  
C;Species: Yersinia pestis  
C;Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001  
C;Accession: AG0161  
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001  
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A;Reference number: AB0001; MUID:21470413; PMID:11586360  
A;Accession: AG0161  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-198 <KUR>  
A;Cross-references: GB:AL590842; PIDN:CAC90154.1; PID:gl5979374; GSPDB:GN00175  
C;Genetics:  
A;Gene: YP01324  
C;Superfamily: bacitracin transport permease; glucose-6-phosphatase catalytic domain hom

Query Match 8.4%; Score 87; DB 2; Length 198;  
Best Local Similarity 32.2%; Pred. No. 0.8;  
Matches 28; Conservative 12; Mismatches 33; Indels 14; Gaps 3;  
Qy 111 ATVLKALIGLDRG-----GECF-----PNVPFPPSDGLEINFEACLYTEG-----BINK 156  
Db 71 ATTCGMLIPDRPVDGFGYAFMSHAPDSGSPSDGHTGTAFTFALAFFWHLHSGVSLMF 130  
Qy 157 LAVNNAFQKMLGIHYRFDGIQGLL 183  
Db 131 IATAIWSRIYLVGHVWPFMDLGAFLG 157

RESULT 4  
F86185  
hypothetical protein [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 17-May-2002  
C;Accession: F86185  
R;Thellogis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizart, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C. C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I. ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: F86185  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-441 <STO>  
A;Cross-references: GB:AE005172; NID:G2388583; PIDN:AAB71464.1; GSPDB:GN00141  
C;Genetics:  
A;Map position: 1  
C;Superfamily: Escherichia coli probable zinc proteinase yael

Query Match 8.3%; Score 86.5; DB 2; Length 441;  
Best Local Similarity 30.7%; Pred. No. 2.6;  
Matches 31; Conservative 11; Mismatches 44; Indels 15; Gaps 4;  
Qy 49 DLDAFDI-----SLLENDELLKRVAEINAAQNPNEVY--LLPQAIQVSGTHSPSPS 101  
Db 147 DNDPDSIPVDDRNLLKRPILDRIVVVSAGIVANVIFAYAIITQVVSGLPVQESFP- 205  
Qy 102 GHATQNGAFATVLKALIGDRGCECPNPVPPSDDGLLEIN 142  
Db 206 -----GVLVPDVKFSASARDG--LLPGDVILAVDGTLSN 239

RESULT 5  
F86920  
probable membrane protein [imported] - Mycobacterium leprae  
C;Species: Mycobacterium leprae  
C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 02-Jun-2003  
C;Accession: F86920  
R;Cole, S.F.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hor R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, eam, M.A.; Rutherford, K.M.  
Nature 409, 1007-1011, 2001  
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq A;Title: Massive gene decay in the leprosy bacillus.  
A;Reference number: A86909; MUID:21128732; PMID:11234002  
A;Accession: F86920  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-192 <STO>  
A;Cross-references: GB:AL450380; NID:gl3092476; PIDN:CAC29602.1; GSPDB:GN00147  
C;Genetics:  
A;Gene: ML0094  
C;Superfamily: bacitracin transport permease; glucose-6-phosphatase catalytic domain hom

Query Match 8.3%; Score 86; DB 2; Length 192;  
Best Local Similarity 24.4%; Pred. No. 0.96;  
Matches 31; Conservative 16; Mismatches 44; Indels 36; Gaps 4;  
Qy 64 LLKRVAEINAAQNPNEVYLLPQAIQVSGTHSPSPGHATQNGAFATVLKALIGLDRG 123  
Db 78 LKRMVR---RIRPNPAV-----TVNVGTPSPSPSAHATSTATAAAILIGRASLPKG 129  
Qy 124 GECFPNPVPPSDDGLLEINFEACLYTEGEINKLVNNAFQKMLGIHYRFDGIQGLL 183  
Db 130 -----IYAIV-----LVAPMALSRIVLVGHVPSDVAFGVVLG 161

Query Match 8.3%; Score 86; DB 2; Length 192;  
Best Local Similarity 24.4%; Pred. No. 0.96;  
Matches 31; Conservative 16; Mismatches 44; Indels 36; Gaps 4;  
Qy 64 LLKRVAEINAAQNPNEVYLLPQAIQVSGTHSPSPGHATQNGAFATVLKALIGLDRG 123  
Db 78 LKRMVR---RIRPNPAV-----TVNVGTPSPSPSAHATSTATAAAILIGRASLPKG 129  
Qy 124 GECFPNPVPPSDDGLLEINFEACLYTEGEINKLVNNAFQKMLGIHYRFDGIQGLL 183  
Db 130 -----IYAIV-----LVAPMALSRIVLVGHVPSDVAFGVVLG 161  
Qy 184 ETITVRT 190  
Db 162 AVAGTT 168  
RESULT 6  
S69334  
chloride peroxidase (EC 1.11.1.10) [validated] - Curvularia inaequalis  
N;Alternate names: vanadium-containing chloroperoxidase  
C;Species: Curvularia inaequalis  
C;Date: 28-Jan-2000 #sequence\_revision 28-Jan-2000 #text\_change 15-Sep-2000  
C;Accession: S69334; S53117

R;Simons, B.H.; Barnett, P.; Vollenbroek, E.G.M.; Dekker, H.L.; Muijsers, A.O.; Messers  
Eur. J. Biochem. 229, 566-574, 1995  
A;Title: Primary structure and characterization of the vanadium chloroperoxidase from th  
A;Reference number: S69334; MUID:95262722; PMID:7744081  
A;Accession: S69334  
A;Molecule type: mRNA  
A;Residues: 1-609 <SIM>  
A;Cross-references: EMBL:X85369; NID:g732617; PIDN:CAA59686.1; PID:g732618  
A;Experimental source: CBS 102.42  
A;Note: part of this sequence was confirmed by sequence analysis of genomic DNA  
R;Messerschmidt, A.; Weyer, R.  
submitted to the Brookhaven Protein Data Bank, September 1995  
A;Reference number: A66871; PDB:1VNC  
A;Contents: annotation; X-ray crystallography, 2.1 angstroms, 3-543, 'E', 545-578  
A;Note: the authors believe that 544-Glu is correct from modeling evidence  
C;Genetics:  
A;Gene: vCPO  
C;Function:  
A;Description: catalyzes the formation of two carbon-chlorine bonds and two water molecu  
C;Superfamily: Curvularia inaequalis chloride peroxidase; glucose-6-phosphatase catalyti  
C;Keywords: blocked amino end; chloride; metalloprotein; oxidoreductase; vanadium  
F;329-511/Domain: Glucose-6-phosphatase catalytic domain homology <GPH>  
F;404/Active site: His #status predicted  
F;496/Binding site: vanadate (His) #status predicted

Query Match 8.2%; Score 85.5; DB 1; Length 609;  
Best Local Similarity 21.7%; Pred. No. 5;  
Matches 30; Conservative 22; Mismatches 35; Indels 51; Gaps 5;  
QY 71 INAAQNPNEVTLLPQAIQVSGTPSPSGHATQNGAFATVLK-----115  
DB 381 LGAPATNTNDIPF-----RPPFPAYPSGHATGGVQFMVRRYNGRVGTWKDDEP 431  
QY 116 -----ALIGDRGCECPNPVPPSDGGLLEL-----NFGAC-LTVEGEINKL 157  
DB 432 DNTAIDMISEELNGVNRDLRQYDPTAPIEDQGVTRIVRHFSQAWELMFENALISRI 491  
QY 158 ANNVAFGRQMLGIHYRFD 175  
DB 492 -----FLGVHWRFD 500

RESULT 7  
D97672  
hypothetical protein AGR\_C\_4732 [imported] - Agrobacterium tumefaciens (strain C58, Cere  
C;Species: Agrobacterium tumefaciens  
C;Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 02-Jun-2003  
C;Accession: D97672  
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A;Reference number: A97359; MUID:21608551; PMID:11743194  
A;Accession: D97672  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-217 <KUR>  
A;Cross-references: GB:AE007869; PIDN:AAK88333.1; PID:g15157811; GSPDB:GN00169  
C;Genetics:  
A;Gene: AGR\_C\_4732  
A;Map position: circular chromosome  
C;Superfamily: bacitracin transport permease; glucose-6-phosphatase catalytic domain hom

Query Match 8.1%; Score 84.5; DB 2; Length 217;  
Best Local Similarity 29.3%; Pred. No. 1.6;  
Matches 27; Conservative 8; Mismatches 28; Indels 29; Gaps 2;  
QY 92 GSPTSPSGHATQNGAFATVLKALIGDRGCECPNPVPPSDGGLINFEACILTYE 151  
DB 116 GDNLYESFPFSGHSTAAGAFGVFAML-----MPRFRWAFLL-- 151  
QY 152 GEINKLVANVAFGRQMLGIHYRFDGQGLLLG 183

Db 152 -----LALVIGSVSRVIVGAHYPSDVAAGLLLG 178

RESULT 8  
AB2897  
conserved hypothetical protein Atu2611 [imported] - Agrobacterium tumefaciens (strain C58  
C;Species: Agrobacterium tumefaciens  
C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 02-Jun-2003  
C;Accession: AB2897  
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A;Reference number: AB2577; MUID:21608550; PMID:11743193  
A;Accession: AB2897  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-242 <KUR>  
A;Cross-references: GB:AE008698; PIDN:AAI43592.1; PID:g17741109; GSPDB:GN00186  
A;Experimental source: strain C58 (Dupont)  
C;Genetics:  
A;Gene: Atu2611  
A;Map position: circular chromosome  
C;Superfamily: bacitracin transport permease; glucose-6-phosphatase catalytic domain hom

Query Match 8.1%; Score 84.5; DB 2; Length 242;  
Best Local Similarity 29.3%; Pred. No. 1.8;  
Matches 27; Conservative 8; Mismatches 28; Indels 29; Gaps 2;  
QY 92 GSPTSPSGHATQNGAFATVLKALIGDRGCECPNPVPPSDGGLINFEACILTYE 151  
DB 141 GDNLYESFPFSGHSTAAGAFGVFAML-----MPRFRWAFLL-- 176  
QY 152 GEINKLVANVAFGRQMLGIHYRFDGQGLLLG 183  
DB 177 -----LALVIGSVSRVIVGAHYPSDVAAGLLLG 203

RESULT 9  
AF2306  
hypothetical protein alr4005 [imported] - Nostoc sp. (strain PCC 7120)  
C;Species: Nostoc sp. PCC 7120  
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C;Accession: AF2306  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
A;Reference number: AB1807; MUID:21595285; PMID:11759840  
A;Accession: AF2306  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-418 <KUR>  
A;Cross-references: GB:BA00019; PIDN:BAW75704.1; PID:g17133140; GSPDB:GN00179  
A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: alr4005

Query Match 8.1%; Score 84; DB 2; Length 418;  
Best Local Similarity 20.8%; Pred. No. 4.2;  
Matches 49; Conservative 28; Mismatches 93; Indels 66; Gaps 9;  
QY 4 GTSHYFRLIGAAELAQASCYQKW-----VHRFA----- 33  
DB 98 GGDQFFVVGKLGKGYRTVYVAEWEARWHNLIDRFVGMKPLVAAKASPKYAHKFTVVVDL 157  
QY 34 -----RPEALGGTLHNTIAGDLADFDLSLENEDELLKRVAEINAQNPNEVTYLLP 86  
DB 158 MLEANAQLPNAPHTPTHTPIIGILPGSKAAKLTOGILPLMUSIGYIHSRKP--ETKEVIP 215

[illegible]

```
Query Match          7.7%; Score 80; DB 2; Length 489;
Best Local Similarity 34.9%; Pred. No. 12;
Matches 30; Conservative 9; Mismatches 39; Indels 8; Gaps 4;

QY 26 KQVHREARPEALGGLTILHTIA-GDLDADEISLLENDELLKRVAEINAAQNPNVEVYL 84
Db 290 KQOHRLLLVNEKLGTTNHRIVACGDVTCGYDLPVANYE--ARTALNNALFLRLQVNYQ 347

QY 85 -LPOAIQVGSPTHPSPYSGHATONGA 109
Db 348 NTPWAIL-----SHPTLAQVGVTEQA 369

RESULT 14
A64822
probable bacitracin transport permease ybjG - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 28-Jan-2000 #sequence_revision 28-Jan-2000 #text_change 01-Mar-2002
C;Accession: A64822
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: A64822
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-198 <BLAT>
A;Cross-references: GB:AE000186; GB:U00096; NID:gl787058; PIDN:AAC73928.1; PID:gl787064;
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: ybjG
C;Superfamily: bacitracin transport permease; glucose-6-phosphatase catalytic domain hom
C;Keywords: antibiotic resistance; transmembrane protein
F;34-50/Domain: transmembrane #status predicted <TM1>
F;51-160/Domain: glucose-6-phosphatase catalytic domain homology <GPH>
F;56-72/Domain: transmembrane #status predicted <TM2>
F;128-144/Domain: transmembrane #status predicted <TM3>
F;151-167/Domain: transmembrane #status predicted <TM4>

Query Match          7.7%; Score 79.5; DB 1; Length 198;
Best Local Similarity 36.8%; Pred. No. 4.2;
Matches 21; Conservative 8; Mismatches 23; Indels 5; Gaps 1;

QY 132 FPSDDGLELINFEGACLTYE-----GEINKLVNVAFGQMLGIHYRFDGIQGLLLG 183
Db 101 FPSDHGTVITTFALAFLCWHLWSGSLLMVLAVVIAWSRVYLVGVHWPDLMLGGLLAG 157

RESULT 15
A99744
hypothetical protein ECs0921 [imported] - Escherichia coli (strain O157:H7, substrain R
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: A99744
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasaawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gend
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: A99744
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-198 <HAY>
A;Cross-references: GB:BA000007; PIDN:BA034344.1; PID:gl3360380; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECe0921
C;Superfamily: bacitracin transport permease; glucose-6-phosphatase catalytic domain hom

Query Match          7.7%; Score 79.5; DB 2; Length 198;
Best Local Similarity 36.8%; Pred. No. 4.2;
```

```
Matches 21; Conservative 8; Mismatches 23; Indels 5; Gaps 1;
QY 132 FPSDDGLELINFEGACLTYE-----GEINKLVNVAFGQMLGIHYRFDGIQGLLLG 183
Db 101 FPSDHGTVITTFALAFLCWHLWSGSLLMVLAVVIAWSRVYLVGVHWPDLMLGGLLAG 157

Search completed: September 19, 2004, 03:59:49
Job time : 13.0069 secs
```

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2004, 03:16:07 ; Search time 36.7002 Seconds  
(without alignments)  
1702.240 Million cell updates/sec

Title: US-10-691-383-2\_COPY\_435\_632

Perfect score: 1039

Sequence: 1 VNFQSHYFLRIGAAELAQ.....GLLGGTITVTLHQELMTF 198

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL\_25:\*
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phase:\*
  - 10: sp\_plant:\*
  - 11: sp\_rhodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_rvirus:\*
  - 16: sp\_bacteriap:\*
  - 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1039	100.0	676	10	O82433
2	503.5	48.5	682	10	Q7X9V0
3	502.5	48.4	646	10	Q7X9V1
4	332.5	32.0	598	10	Q8LLW7
5	331.5	31.9	598	10	O81959
6	322	31.0	597	10	O81960
7	116.5	11.2	145	16	Q9RVL3
8	108	10.4	165	16	O53584
9	108	10.4	165	16	Q7TVM5
10	104	10.0	416	16	Q89ER7
11	99	9.5	706	16	Q7UUV2
12	93	9.0	525	16	Q825G8
13	92.5	8.9	299	10	Q94AY8
14	90.5	8.7	441	10	Q8L738
15	90.5	8.7	667	13	Q7T2E5
16	88.5	8.5	396	5	Q9VDN3

17	88.5	8.5	647	4	Q9H6T8
18	87	8.4	198	16	Q8ZGH3
19	87	8.4	201	16	Q8D031
20	87	8.4	308	2	Q9RPH0
21	86.5	8.3	248	3	P78610
22	86.5	8.3	441	10	O23053
23	86.5	8.3	487	2	P72262
24	86	8.3	192	16	Q9CDB6
25	86	8.3	201	17	Q8TW45
26	85.5	8.2	367	11	Q8R2X1
27	85.5	8.2	881	11	Q8C0E7
28	84.5	8.1	242	16	Q8UC87
29	84.5	8.1	289	16	Q7TUQ8
30	84.5	8.1	1431	10	Q7XR16
31	84	8.1	168	16	Q8NLQ8
32	84	8.1	307	16	Q8FLZ6
33	84	8.1	418	16	Q8YC31
34	84	8.1	453	16	Q82PX4
35	84	8.1	612	16	Q88VB9
36	83.5	8.0	626	4	Q9NXC4
37	83	8.0	271	16	Q98GT9
38	83	8.0	421	16	Q7UHC0
39	83	8.0	465	16	Q89AP5
40	83	8.0	800	13	Q918X3
41	82	7.9	308	2	Q9RFH7
42	82	7.9	308	2	Q9RFH2
43	81.5	7.8	235	16	Q92KY7
44	81.5	7.8	487	2	P72269
45	81.5	7.8	619	2	Q8VTN2

ALIGNMENTS

RESULT 1

O82433	PRELIMINARY;	PRT;	676 AA.
ID	O82433		
AC	O82433;		
DT	01-NOV-1998 (TrEMBLrel. 08, Created)		
DT	01-NOV-1998 (TrEMBLrel. 08, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	Vanadium bromoperoxidase.		
OS	Fucus distichus.		
OC	Eukaryota; stramenopiles; Phaeophyceae; Fucales; Fucaceae; Fucus.		
OX	NCBI_TaxID=3012;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Vreeland V., Ng K., Epstein L.;		
RL	Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF053411; AAC35279.1; -		
DR	HSSP; P81701; 1Q19.		
DR	GO; GO:0004601; F:peroxidase activity; IEA.		
DR	InterPro; IPR008934; ACPase_VanPerase.		
DR	InterPro; IPR000326; FA_PTPase.		
KW	Peroxidase.		
SQ	SEQUENCE 676 AA; 73344 MW; AA8A46F539A14EC7 CRC64;		

Query Match	100.0%;	Score 1039;	DB 10;	Length 676;
Best Local Similarity	100.0%;	Pred. No. 1.8e-90;		
Matches 198;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	VNFQSHYFLRIGAAELAQASCYQKWVHRFARPEALGGLTHTNTIAGDLADFDISLLE	60	
DB	435	VNFQSHYFLRIGAAELAQASCYQKWVHRFARPEALGGLTHTNTIAGDLADFDISLLE	494	
QY	61	NDELLKRVABINAAQPNNEVTYLLPOAIOVGSPHPSYSGHATQNGAFATVLKALIGL	120	
DB	495	NDELLKRVABINAAQPNNEVTYLLPOAIOVGSPHPSYSGHATQNGAFATVLKALIGL	554	
QY	121	DRGECFPNPFVSDGGLLEINFEAGLTYEGEINKLVANVAFGRQMLGTHYRFDGIQGL	180	
DB	555	DRGECFPNPFVSDGGLLEINFEAGLTYEGEINKLVANVAFGRQMLGTHYRFDGIQGL	614	

```
QY 181 LLGETITVTRLHQLMTF 198
DB 615 LLGETITVTRLHQLMTF 632

RESULT 2
Q7X9V0 PRELIMINARY; PRT; 682 AA.
AC Q7X9V0;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Vanadium-dependent bromoperoxidase 2.
GN VBPO2.
OS Laminaria digitata.
OC Eukaryota; stramenopiles; Phaeophyceae; Laminariales; Laminariaceae;
OC Laminaria.
OX NCBI_TaxID=80365;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=sporophyte;
RX MEDLINE=22703831; PubMed=12697758;
RA Colin C., Leblanc C., Wagner E., Delage L., Leize-Wagner E.,
RA van Dorselaer A., Kloareg B., Potin P.;
RT "The Brown Algal Kelp Laminaria digitata Features Distinct
RT Bromoperoxidase and Iodoperoxidase Activities.";
RL J. Biol. Chem. 278:23545-23552(2003).
DR EMBL; AJ491787; CAD37192.1; -.
KW Peroxidase.
SQ SEQUENCE 682 AA; 75125 MW; 98DD4FC45B17E772 CRC64;

Query Match 48.5%; Score 503.5; DB 10; Length 682;
Best Local Similarity 46.7%; Pred. No. 2.8e-39;
Matches 98; Conservative 28; Mismatches 67; Indels 17; Gaps 2;

QY 3 FGTSHYFRLLIGAAELAQASCYQKQVHRPAREALGGTLHNTIAGDLADDFDISLLEND 62
DB 386 FGEPHILTAMASASSSTRHAWYAKQVHRMLRPEAYGALVHNTLMRDVITPLPDSILRNT 445

QY 63 ELLKRVAEINAAONPNNEVYLLPQAIQVGSPTSPHPSGHATONGAFATVTKALIGLDR 122
DB 446 ELLNRVEVHNRQNPDPGEKTFLLPMAAQQSPHPAYPSGHAINNGAYITALKAFILGYEA 505

QY 123 GGCFPNVPFPPSDGGLLEINFE-----GAC-----LTYEGEINKLAVNVAFGR 165
DB 506 GQKCFPNVVSNDGDKRIKYPGSGREIVGECVNEKGKLVGELYEGELNKISANVLLGR 565

QY 166 QMLGIHYRFDGIQGLLIGETITVTRLHQL 195
DB 566 SHIGVHRMDGVYGALMGETSCVRRLOQEL 595

RESULT 3
Q7X9V1 PRELIMINARY; PRT; 646 AA.
AC Q7X9V1;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Vanadium-dependent bromoperoxidase 1.
GN VBPO1.
OS Laminaria digitata.
OC Eukaryota; stramenopiles; Phaeophyceae; Laminariales; Laminariaceae;
OC Laminaria.
OX NCBI_TaxID=80365;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=sporophyte;
RX MEDLINE=22703831; PubMed=12697758;
RA Colin C., Leblanc C., Wagner E., Delage L., Leize-Wagner E.,
RA van Dorselaer A., Kloareg B., Potin P.;
RT "The Brown Algal Kelp Laminaria digitata Features Distinct
RT Bromoperoxidase and Iodoperoxidase Activities.";
```

```
RL J. Biol. Chem. 278:23545-23552(2003).
DR EMBL; AJ491786; CAD37191.1; -.
KW Peroxidase.
SQ SEQUENCE 646 AA; 71171 MW; DB9382F74CBB97FD CRC64;

Query Match 48.4%; Score 502.5; DB 10; Length 646;
Best Local Similarity 46.7%; Pred. No. 3.2e-39;
Matches 98; Conservative 27; Mismatches 68; Indels 17; Gaps 2;

QY 3 FGTSHYFRLLIGAAELAQASCYQKQVHRPAREALGGTLHNTIAGDLADDFDISLLEND 62
DB 350 FGEPHILTAMASASSSTRHAWYAKQVHRMLRPEAYGALVHNTLMRDVITPLPDSILRNT 409

QY 63 ELLKRVAEINAAONPNNEVYLLPQAIQVGSPTSPHPSGHATONGAFATVTKALIGLDR 122
DB 410 ELLNRVEVHNRQNPDPGEKTFLLPMAAQQSPHPAYPSGHAINNGAYITALKAFILGYEA 469

QY 123 GGCFPNVPFPPSDGGLLEINFE-----GAC-----LTYEGEINKLAVNVAFGR 165
DB 470 GQKCFPNVVSNDGDKRIKYPGSGREIVGECVNEKGKLVGELYEGELNKISANVLLGR 529

QY 166 QMLGIHYRFDGIQGLLIGETITVTRLHQL 195
DB 530 SHIGVHRMDGVYGALMGETSCVRRLOQEL 559

RESULT 4
Q8LLW7 PRELIMINARY; PRT; 598 AA.
AC Q8LLW7;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Vanadium-dependent bromoperoxidase.
OS Corallina officinalis (Coral seaweed).
OC Eukaryota; Rhodophyta; Florideophyceae; Corallinales; Corallinaceae;
OC Corallinoideae; Corallina.
OX NCBI_TaxID=35170;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22117781; PubMed=12121762;
RA Carter J.N., Beatty K.E., Simpson M.T., Butler A.;
RT "Reactivity of recombinant and mutant vanadium bromoperoxidase from
RT the red alga Corallina officinalis.";
RL J. Inorg. Biochem. 91:59-69(2002).
DR EMBL; AF218810; AM46061.1; -.
DR PDB; 1QHB; 05-JUL-00.
DR GO; GO:0004601; P:peroxidase activity; IEA.
DR InterPro; IPR008934; AcPase VanPerase.
DR InterPro; IPR000326; PA_PTPase.
KW Peroxidase.
SQ SEQUENCE 598 AA; 65458 MW; BC784E370D748F01 CRC64;

Query Match 32.0%; Score 332.5; DB 10; Length 598;
Best Local Similarity 38.5%; Pred. No. 5.1e-23;
Matches 82; Conservative 38; Mismatches 72; Indels 21; Gaps 7;

QY 1 VNEGTSHYFRLLIGAAELAQASCYQKQVHRPAREALGGTLH-NTIAGD----LDADF 54
DB 374 VNFGDHVLVLSTVEATRALKAVRYQKFNHRLRPEATGGLISVNKIAAEKGESVPEV 433

QY 55 DISLLENDELKRVAEINAAQN-----PNNEVYLLPQAIQVGSPTSPHPSGHATONGA 109
DB 434 DLAVEELEDLEKAEISNRKQNTADGDDPDPDFLLPQAFAGSPHPSPSGSHAVAGA 493

QY 110 FATVLKALIG-----LDRGGCEFPNVPFSDGGLLEINFEACITYPEGEINKLAVNVAFGR 165
DB 494 CVTILKAFDPSNFQIDQVFE-----VDKDBDKLVKSFSGT-LTVAGELNKLADNIAIGR 547

QY 166 QMLGIHYRFDGIQGLLIGETITVTRLHQLMTF 198
DB 548 NMAGVHYFSDQFESILLGQVQVATGILLESRLTY 580
```





```
Db 49 EYNVAR-POSMBDHVQGMWAPSLPTPPSPVPSGHATVSGAAAEVL-----AQFPP 98
QY 129 NPVPSPDDGLELINFEGACTYGEINKLVAVNFAFGOMLGIHYRFGIOGLLGETI 186
Db 99 -----LQARQLRRDARDAAFSRVGGIHWGVDGVALGDVGQRV 136

RESULT 8
OS3584 PRELIMINARY; PRT; 165 AA.
AC OS3584;
DT 01-JUN-1998 (T-EMBLrel. 06, Created)
DT 01-JUN-1998 (T-EMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Hypothetical protein (PAP2 superfamily protein).
GN RV3807C OR MTV026.12 OR MT3914.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=9825987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagals K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne K., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Unayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022076; CAAL17871.1; -.
DR EMBL; AE007185; AAK48280.1; -.
DR PIR; C70888; C70888.
DR TIGR; MT3914; -.
DR TubercuList; RV3807c; -.
DR InterPro; IPR008934; AcPase_VanPerase.
DR InterPro; IPR000326; PA_PTPase.
DR Pfam; PF01569; PAP2; 1.
DR SMART; SM00014; acidPPc; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 165 AA; 17218 MW; 0E084FA26508A4A1 CRC64;

Query Match 10.4%; Score 108; DB 16; Length 165;
Best Local Similarity 24.0%; Pred. No. 0.025;
Matches 46; Conservative 22; Mismatches 52; Indels 72; Gaps 8;

QY 4 GTSHY-----FRLGAAELAQASCVQKQVHRFARPEALGGTLHNTIAGDLDAF 54
Db 20 GLSHFGHCIGWLILALLGIALPRR---REMLV-----AGAGAFVAHAIA----- 63
QY 55 DISLENDELLKRVAEINAAQNPNEVYLLPOAIOVGSPTSPSGHATONGAFATVL 114
Db 64 -----VLTKLVR---RQPDHPAI-----AVNVDTSPQLSPSAHATSTTAAALLM 107
QY 115 KALIGLDRGCECPNPVPFSDGGLLEINFEGACTYGEINKLVAVNFAFGOMLGIHYRF 174
Db 108 GRATGL-----PLPVV-----LVPPMALSRIILGVHYPYS 136

QY 175 DGIQGLLGETI 186
Db 137 DVAVGVALGATV 148

RESULT 10
Q89ER7 PRELIMINARY; PRT; 416 AA.
AC Q89ER7;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE BLr7005 protein.
GN BLR7005.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
```

```
QY 175 DGIQGLLGETI 186
Db 137 DVAVGVALGATV 148

RESULT 9
Q7TW5 PRELIMINARY; PRT; 165 AA.
AC Q7TW5;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Possible conserved transmembrane protein.
GN MB3837C.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pyor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248347; CAD96023.1; -.
KW Complete proteome.
SQ SEQUENCE 165 AA; 17218 MW; 0E084FA26508A4A1 CRC64;

Query Match 10.4%; Score 108; DB 16; Length 165;
Best Local Similarity 24.0%; Pred. No. 0.025;
Matches 46; Conservative 22; Mismatches 52; Indels 72; Gaps 8;

QY 4 GTSHY-----FRLGAAELAQASCVQKQVHRFARPEALGGTLHNTIAGDLDAF 54
Db 20 GLSHFGHCIGWLILALLGIALPRR---REMLV-----AGAGAFVAHAIA----- 63
QY 55 DISLENDELLKRVAEINAAQNPNEVYLLPOAIOVGSPTSPSGHATONGAFATVL 114
Db 64 -----VLTKLVR---RQPDHPAI-----AVNVDTSPQLSPSAHATSTTAAALLM 107
QY 115 KALIGLDRGCECPNPVPFSDGGLLEINFEGACTYGEINKLVAVNFAFGOMLGIHYRF 174
Db 108 GRATGL-----PLPVV-----LVPPMALSRIILGVHYPYS 136

QY 175 DGIQGLLGETI 186
Db 137 DVAVGVALGATV 148

RESULT 10
Q89ER7 PRELIMINARY; PRT; 416 AA.
AC Q89ER7;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE BLr7005 protein.
GN BLR7005.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
```

```
RA Kohara M., Matsumoto M., Shimpō S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RL Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR ENBL; AP005960; BAC52270.1; -.
GO; GO:0004601; F:peroxidase activity; IEA.
DR InterPro; IPR008934; ACPase_VanPerase.
DR InterPro; IPR003320; Cl_Perox.
DR Pfam; PF02328; Cl_Perox; 1.
KW Complete proteome.
SQ SEQUENCE 416 AA; 4988 MW; F01F13A3E31051C CRC64;

Query Match 10.0%; Score 104; DB 16; Length 416;
Best Local Similarity 27.1%; Pred. No. 0.21;
Matches 42; Conservative 23; Mismatches 70; Indels 20; Gaps 6;

QY 45 TIAGD--LDADFDISLLENDELKRVAEINAAQNPNNEVYLLPQAIQVG-SPTHSYPS 101
Db DNA Res. 9:189-197(2002).
267 SIAGNDALVAVFAKHYNN--LWRPITAIRNADLTSPATPRDPFWLPLGETPMHPEYPC 324
QY 102 GHATQNGAFATVLKALIGLDRGCECPNVPFSDGGLINFEAGCLTYE-GBINKLAVN 160
Db ACITSAATSTVLQSVVG-DGFEFSLTSSAPG-----VTRKWSRLQDYSDE 370
QY 161 VAFGRQMLGHYRFDGIGQLLGLGETITVRTLHQL 195
Db VSNARIWAGHYRFEVSGDKMGKIGALTVAQL 405

RESULT 11
Q7UVW2 PRELIMINARY; PRT; 706 AA.
AC Q7UVW2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Vanadium chloroperoxidase (EC 1.11.1.10).
GN CPO OR RB2401.
OS Rhodospirillum rubrum.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1; PubMed=12835416;
RX Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA Schleutner H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
strain 1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
DR ENBL; BX294137; CAD72609.1; -.
KW Oxidoreductase; Peroxidase; Complete proteome.
SQ SEQUENCE 706 AA; 76527 MW; 9512A7A54206A6F8 CRC64;

Query Match 9.5%; Score 99; DB 16; Length 706;
Best Local Similarity 28.1%; Pred. No. 1.3;
Matches 45; Conservative 18; Mismatches 67; Indels 30; Gaps 8;

QY 39 GGTLLHTIAGLDADFD--ISLLENDELKRVAEINAAQNPN--NEVYLLPQAIQVGSPT 95
Db GEELRPVTAIQSGDFDGNLTGSD-----ADWTALGAPDGGDIDVGFTFPQ----- 517
QY 96 HPSYPGSHATQNGAFATVLKALIGLDRGCECPNVPFSDGGLINFEAGCLTYE-GBINKLAVN 146
Db FTYISGHATFGGALFGLTQFEYCTDDISFTVASEEILLDNPELQEAAYGL---NLDDA 574
QY 147 CLTYEGEINKLAVNVAFRQMLGHYRFDGIGQLLGLGETI 196
Db ERTFSFSEAWAEN-GRSRYVLGLIHFDLVDVGQVQSI 613
```

## RESULT 12

```
Q825G8 PRELIMINARY; PRT; 525 AA.
ID Q825G8;
AC Q825G8;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN SAV7489.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteriae; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
DR ENBL; AP005050; BAC75200.1; -.
GO; GO:0004143; F:diacylglycerol kinase activity; IEA.
DR GO; GO:0007205; P:protein kinase C activation; IEA.
DR InterPro; IPR008934; ACPase_VanPerase.
DR InterPro; IPR001206; DAGKC.
DR InterPro; IPR000326; PA_PTPase.
DR Pfam; PF00781; DAGKC; 1.
DR Pfam; PF01569; PAP2; 1.
DR SMART; SM00014; acidppc; 1.
DR SMART; SM00046; DAGKC; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 525 AA; 55271 MW; 38C21722F7996F17 CRC64;

Query Match 9.0%; Score 93; DB 16; Length 525;
Best Local Similarity 27.7%; Pred. No. 3.3;
Matches 28; Conservative 12; Mismatches 31; Indels 30; Gaps 3;

QY 84 LLPQAIQVGSPTH-PSYPSGHATQNGAFATVLKALIGLDRGCECPNVPFSDGGLIN 142
Db LVPQIRHLTRQPHSTTSPGSHSASAAAFATG-----VALESTG----- 127
QY 143 FEGACLTYEGEINKLAVNVAFRQMLGHYRFDGIGQLLGL 183
Db -----YCALVAPLAAAVAFSRVYVGVHYPGVDLGMAIG 161

RESULT 13
Q94AY8 PRELIMINARY; PRT; 299 AA.
ID Q94AY8;
AC Q94AY8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE AC1905140/YUP812.25.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
```



```
Qy 97 -PSYPSGHA-TONGAFATVTLKALIGLDRGGECFENPVFPSPD--DGLBLINFE----- 144
Db 62 ISAEFQTHQDLNQSFRRVAKS-----DRSKDLIKSAILDNDFMKNLEWSQIQEIVDCWYP 117
Qy 145 -----GACLTVEGEINKLAVNVAEGROMLGIHYRFDGIQ-----GLLLGE----- 184
Db 118 VDYDKNSCIIKEBGDVGSLVYVMEDEK----VEVTKEGLKLCMTMGPGKVGELAILYNCTR 173
Qy 185 TITVRTL 191
Db 174 TATVRTV 180
```

Search completed: September 19, 2004, 03:58:50  
Job time : 39.7002 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.  
OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 19, 2004, 03:59:53 ; Search time 1594.65 Seconds  
(without alignments)  
3707.849 Million cell updates/sec

Title: US-10-691-383-2\_COPY\_435\_632  
Perfect score: 1039  
Sequence: 1 VNFQTHYFLRIGAAELAQ.....GLLGLTITVRLHQLMTF 198  
Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Zgapop 6.0 , Zgapext 7.0  
Delop 6.0 , Delext 7.0  
Searched: 27513289 seqs, 14931090276 residues  
Total number of hits satisfying chosen parameters: 55026578  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2 1/USPTO.spool/US10691383/runat.17092004.102656.1753/app.query.fasta\_1.1230  
-DB=EST -QPMT=fastcap -SUFFIX=rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bites -START=1 -END=1 -MATRIX=dlosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10691383 @CGN 1 1 4591 @runat.17092004.102656.1753 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_esti:\*  
10: gb\_estc2:\*  
11: gb\_hic:\*  
12: gb\_estc3:\*  
13: gb\_estc4:\*  
14: gb\_estc5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_man:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	177.5	17.1	503	9	AU187248	AU187248 AU187248
2	177.5	17.1	538	9	AU191942	AU191942 AU191942
3	113	10.9	369	14	CD063088	CD063088 MA1-0029U
4	94.5	9.1	794	12	BI761471	BI761471 603046604
5	93.5	9.0	558	10	AW661332	AW661332 833004D10
6	93	9.0	1207	28	BZ579009	BZ579009 msh2.6117
7	92.5	8.9	745	14	CB037634	CB037634 TC_ad2.36
8	92	8.9	818	14	CB977242	CB977242 CAB400003
9	91.5	8.8	537	14	CF192410	CF192410 14f06j2.E
C 10	90.5	8.7	615	14	CA239663	CA239663 SCBFL506
11	88.5	8.5	604	14	CB925226	CB925226 ABA1.20.E
12	88.5	8.5	612	13	BQ810389	BQ810389 1030017B0
13	88.5	8.5	629	14	CB331374	CB331374 3529.1.36
14	88.5	8.5	639	14	CD223063	CD223063 CCC1_25.C
15	88.5	8.5	666	14	CD229795	CD229795 CCC1_19.D
16	88.5	8.5	755	12	BI772791	BI772791 603054341
C 17	87.5	8.4	743	14	CF571423	CF571423 MCS013F01
18	87.5	8.4	780	14	CK025517	CK025517 AGENCOURT
C 19	87.5	8.4	944	28	CC327659	CC327659 OGUPEP5TH
20	87	8.4	599	14	CF485754	CF485754 POL1_33.A
21	87	8.4	703	12	BJ272534	BJ272534 BJ272534
22	86.5	8.3	468	9	AL045228	AL045228 DKF2p434H
23	86.5	8.3	577	13	BQ829915	BQ829915 L6in2239
C 24	86.5	8.3	688	12	BJ598040	BJ598040 BJ598040
C 25	86.5	8.3	718	12	BJ600919	BJ600919 BJ600919
26	86	8.3	710	13	BW208048	BW208048 BW208048
C 27	86	8.3	3218	11	AK088408	AK088408 Mus muscu
C 28	86	8.3	3261	29	AY408110	AY408110 Mus muscu
C 29	86	8.3	4324	11	AK085707	AK085707 Mus muscu
C 30	85.5	8.2	568	13	BQ811943	BQ811943 L6in1122
31	85.5	8.2	592	10	BE382065	BE382065 601272225
C 32	85.5	8.2	620	28	BZ332752	BZ332752 hx30d08.b
33	85.5	8.2	629	10	BE374596	BE374596 601225384
34	85.5	8.2	726	12	BI648082	BI648082 603279290
C 35	85.5	8.2	744	12	BJ598045	BJ598045 BJ598045
36	85.5	8.2	773	12	BI157032	BI157032 602921306
37	85.5	8.2	785	12	BI688650	BI688650 603311262
38	85.5	8.2	787	12	BI156730	BI156730 602922150
39	85.5	8.2	844	12	BI158626	BI158626 602922395
40	85.5	8.2	887	14	CA980253	CA980253 AGENCOURT
C 41	85.5	8.2	903	10	BF311105	BF311105 601898555
42	85.5	8.2	1065	12	BM924437	BM924437 AGENCOURT
43	85.5	8.2	1988	11	BC023725	BC023725 Mus muscu
44	85.5	8.2	3110	11	AK031542	AK031542 Mus muscu
45	85.5	8.2	3372	11	BC046553	BC046553 Mus muscu

ALIGNMENTS

RESULT 1  
AU187248  
LOCUS  
DEFINITION AU187248 Porphyra yezoensis TU-1 sporophytes Porphyra yezoensis  
CDNA clone PF010b07\_r 5', mRNA sequence.  
EST.  
ACCESSION AU187248  
VERSION AU187248.1 GI:31920687  
KEYWORDS  
SOURCE Porphyra yezoensis  
ORGANISM Porphyra yezoensis  
Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;  
Porphyra.  
REFERENCE 1 (bases 1 to 503)

AUTHORS Asamizu,E., Nakajima,M., Kitade,Y., Saga,N., Nakamura,Y. and Tabata,S.  
TITLE COMPARISON OF RNA EXPRESSION PROFILES BETWEEN THE TWO GENERATIONS OF PORPHYRA YEZOENSIS (RHODOPHYTA), BASED ON EXPRESSED SEQUENCE TAG FREQUENCY ANALYSIS  
JOURNAL J. Phycol. 39 (5), 923-930 (2003)  
COMMENT Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.  
FEATURES  
Location/Qualifiers  
source  
1. 503  
/organism="Porphyra yezoensis"  
/mol\_type="mRNA"  
/strain="TU-1"  
/db\_xref="taxon:2788"  
/clone="PF010b07\_r"  
/dev\_stage="sporophytes"  
/clone\_lib="Porphyra yezoensis TU-1 sporophytes"

ORIGIN

Alignment Scores:  
Pred. No.: 2,21e-10 Length: 503  
Score: 177.50 Matches: 53  
Percent Similarity: 51.64% Conservative: 10  
Best Local Similarity: 43.44% Mismatches: 50  
Query Match: 17.08% Indels: 9  
DB: 4

US-10-691-383-2\_COPY\_435\_632 (1-198) x AU187248 (1-503)

Qy 3 PheGlyThrSerHisTyrPheArgLeuIle---GlyAlaAlaGluLeuAlaGlnArgAla 21  
Db 96 TTTGGGGGGCCCCCACCCTGCTGACGCTGTCACCGAGGTGGCCACACGCGCCTCAAGGCG 155  
Qy 22 SerCysTyrGlnIlystrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThr 41  
Db 156 GTGGCTTCCAAAAGTTCAACGTTCCACGCGCGCTCCGGCCCGAGGGGGTGGGGGCTC 215  
Qy 42 LeuHisAsn---ThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGlu 60  
Db 216 CTCGACCGCTGCGCGCGCCAGGACCGCTGCGCGCAGTGGCGCCCGCTGGCAGAGGAG 275  
Qy 61 AsnAspGluLeuLeuIlysArgValAlaGluLeuAsnAlaIle-----Gln 75  
Db 276 GCGCGCAGACGCTCCCGCGGTGGTGTGATGCCAACGCTGCCCGCTGCGGGGGCGAGT 335  
Qy 76 AsnProAsnAsnGluValThrTyr-----LeuLeuProGlnAlaIleGlnValGlySer 93  
Db 336 GCGCGCGGTGGCGCGCTTACGACAGTGTGCTGCTGCAATGGCTTCCAGAGGGCTCC 395  
Qy 94 ProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrVal 113  
Db 396 CCGATGACCCCTCTCTACGCGCGCGGTACGCCACCGTGGCGGGCGGTGCGTGACGGTG 455  
Qy 114 LeuIlys 115  
Db 456 CTCAAG 461

RESULT 2  
AU191942  
LOCUS AU191942 Porphyra yezoensis TU-1 sporophytes Porphyra yezoensis  
DEFINITION cDNA clone PFL023f08\_r 5', mRNA sequence.  
ACCESSION AU191942  
VERSION AU191942  
KEYWORDS AU191942.1 GI:31930088  
SOURCE EST.  
ORGANISM Porphyra yezoensis  
Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;  
Porphyra.  
REFERENCE 1 (bases 1 to 538)

AUTHORS Asamizu,E., Nakajima,M., Kitade,Y., Saga,N., Nakamura,Y. and Tabata,S.  
TITLE COMPARISON OF RNA EXPRESSION PROFILES BETWEEN THE TWO GENERATIONS OF PORPHYRA YEZOENSIS (RHODOPHYTA), BASED ON EXPRESSED SEQUENCE TAG FREQUENCY ANALYSIS  
JOURNAL J. Phycol. 39 (5), 923-930 (2003)  
COMMENT Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.  
FEATURES  
Location/Qualifiers  
source  
1. 538  
/organism="Porphyra yezoensis"  
/mol\_type="mRNA"  
/strain="TU-1"  
/db\_xref="taxon:2788"  
/clone="PFL023f08\_r"  
/dev\_stage="sporophytes"  
/clone\_lib="Porphyra yezoensis TU-1 sporophytes"

ORIGIN

Alignment Scores:  
Pred. No.: 2,47e-10 Length: 538  
Score: 177.50 Matches: 53  
Percent Similarity: 51.64% Conservative: 10  
Best Local Similarity: 43.44% Mismatches: 50  
Query Match: 17.08% Indels: 9  
DB: 4

US-10-691-383-2\_COPY\_435\_632 (1-198) x AU191942 (1-538)

Qy 3 PheGlyThrSerHisTyrPheArgLeuIle---GlyAlaAlaGluLeuAlaGlnArgAla 21  
Db 115 TTTGGGGGGCCCCCACCCTGCTGACGCTGTCACCGAGGTGGCCACACGCGCCTCAAGGCG 174  
Qy 22 SerCysTyrGlnIlystrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThr 41  
Db 175 GTGGCTTCCAAAAGTTCAACGTTCCACGCGCGCTCCGGCCCGAGGGGGTGGGGGCTC 234  
Qy 42 LeuHisAsn---ThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGlu 60  
Db 235 CTCGACCGCTGCGCGCGCCAGGACCGCTGCGCGCAGTGGCGCCCGCTGGCAGAGGAG 294  
Qy 61 AsnAspGluLeuLeuIlysArgValAlaGluLeuAsnAlaIle-----Gln 75  
Db 295 GCGCGCAGACGCTCCCGCGGTGGTGTGATGCCAACGCTGCCCGCTGCGGGGGCGAGT 354  
Qy 76 AsnProAsnAsnGluValThrTyr-----LeuLeuProGlnAlaIleGlnValGlySer 93  
Db 355 GCGCGCGGTGGCGCGCTTACGACAGTGTGCTGCTGCAATGGCTTCCAGAGGGCTCC 414  
Qy 94 ProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrVal 113  
Db 415 CCGATGACCCCTCTCTACGCGCGCGGTACGCCACCGTGGCGGGCGGTGCGTGACGGTG 474  
Qy 114 LeuIlys 115  
Db 475 CTCAAG 480

RESULT 3  
CD063088  
LOCUS CD063088  
DEFINITION MA1-0029U-M019-D02-U.B MA1-0029 Schistosoma mansoni cDNA clone  
ACCESSION MA1-0029U-M019-D02.B, mRNA sequence.  
VERSION CD063088  
KEYWORDS CD063088.1 GI:34614247  
SOURCE EST.  
ORGANISM Schistosoma mansoni  
Schistosoma mansoni  
Schistosoma mansoni  
Schistosoma mansoni  
Strigeidida; Schistosomatidae; Schistosoma.  
REFERENCE 1 (bases 1 to 369)



**AUTHORS** Verjowski-Almeida, S., DeMarco, R., Martins, E.A.L., Guimaraes, P.E.M., Ojopi, E.P.B., Paquola, A.C.M., Piazza, J.P., Nishiyama, M.Y. Jr., Kitajima, J.P., Adams, R.E., Ashton, P.D., Bonaldo, M.F., Coulson, P.S., Dillon, G.P., Farias, L.P., Gregorio, S.P., Ho, P.L., Leite, R.A., Malaquias, L.C.C., Marques, R.C.P., Miyasato, P.A., Nascimento, A.L.T.O., Ohlweiler, F.P., Reis, E.M., Ribeiro, M.A., Sa, R.G., Stukart, G.C., Soares, M.B., Gargioni, C., Kawano, T., Rodrigues, V., Madeira, A.M.B.N., Wilson, R.A., Menck, C.F.M., Setubal, J.C., Leite, L.C.C. and Dias-Neto, E.

**TITLE** Transcriptome analysis of the acelomate human parasite Schistosoma mansoni

**JOURNAL** Nat. Genet. 35 (2), 148-157 (2003)

**MEDLINE** 22879926

**COMMENT** Contact: Dr. Sergio Verjowski-Almeida  
Departamento de Bioquímica  
Instituto de Química - Universidade de São Paulo  
Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP, Brasil  
Tel: +55-11-3091-2173  
Fax: +55-11-3091-2186  
Email: verjo@iq.usp.br

This sequence was derived from the FAPESP Schistosoma mansoni EST Genome Project. All sequences in the project were assembled and annotated. This entry and all the assembled sequences can be seen in the following URL <http://bioinfo.iq.usp.br/schisto/>

Plate: MA1-0029U-M019 row: 2 column: D.

**FEATURES**

Location/Qualifiers	1..369
source	1..369
organism="Schistosoma mansoni"	
mol_type="mRNA"	
db_xref="taxon:6183"	
clone="MA1-0029U-M019-D02.B"	
sex="mixed pool"	
dev stage="adult"	
lab_host="Mesocricetus auratus"	
clone_lib="MA1-0029"	
note="vector: SureClone"	

**ORIGIN**

Alignment Scores:

Pred. No.:	0.00515	Length:	369
Score:	113.00	Matches:	29
Percent Similarity:	47.52%	Conservative:	19
Best Local Similarity:	28.71%	Mismatches:	33
Query Match:	10.88%	Indels:	20
DB:	14	Gaps:	3

US-10-691-383-2\_COPY\_435\_632 (1-198) x CD063088 (1-369)

Qy	86	ProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThr	105
Db	1	CCGAATAGCTCGATTATCCGACGCGCGCATGGCACATTCCGCGGCGCACGCGTG	60
Qy	106	GlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlu	125
Db	61	GAAGCATTCGCGATGACGCGAGGTCTCAAGCGGCTCATA-----GAT	102
Qy	126	CysPheProAsnProValPheProSerAspAspGlyLeuGluLeuIleAsnPheGluGly	145
Db	103	TGCGTGGCGCAT-----TATCCGATATACAAAGCGCGCATGCTCG-----	144
Qy	146	AlaCysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArg	165
Db	145	-----CTGGACAGCTCGCGGCGCATCGCATACACCG	180
Qy	166	GlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThr	185
Db	181	GTGGTTCGCGGGTACACTTTCGATCGACATCTGGCGGGGTGCACTGCTTGGACGCTCG	240
Qy	186	Ile	186
Db	241	ATC	243

**RESULT 4**

BI761471

**LOCUS** 794 bp mRNA linear EST 25-SEP-2001

**DEFINITION** 603046604F1 NIH\_MGC\_116 Homo sapiens CDNA clone IMAGE:5186736 5', mRNA sequence.

**ACCESSION** BI761471

**VERSION** BI761471.1 GI:15753049

**KEYWORDS** EST.

**SOURCE** Homo sapiens (human)

**ORGANISM** Homo sapiens

**REFERENCE** 1 (bases 1 to 794)

**AUTHORS** NIH-MGC <http://mgc.nci.nih.gov/>.

**TITLE** National Institutes of Health, Mammalian Gene Collection (MGC)

**JOURNAL** Unpublished (1999)

**COMMENT** Unpublished: Robert Strauberg, Ph.D.  
Email: csapbs@email.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LLAM11466 row: h column: 01  
High quality sequence stop: 741.

**FEATURES**

Location/Qualifiers	1..794
source	1..794
organism="Homo sapiens"	
mol_type="mRNA"	
db_xref="taxon:9606"	
clone="IMAGE:5186736"	
lab_host="DH10B"	
clone_lib="NIH_MGC_116"	
note="Organ: pooled colon, kidney, stomach; Vector: pCMV-SPORT6; Site: 1: NotI; Site: 2: EcoRV (destroyed); RNA source anonymous pool of 3 colons, age 26 yo male, 49 yo female, 71 yo male colon; 46 yo male kidney, and pool of 2 stomachs, 62 yo male and 70 yo female. Library is oligo-dT primed and directionally cloned [EcoRV site is destroyed upon cloning]. Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIH_MGC Library."	

**ORIGIN**

Alignment Scores:

Pred. No.:	2.74	Length:	794
Score:	94.50	Matches:	39
Percent Similarity:	42.54%	Conservative:	18
Best Local Similarity:	29.10%	Mismatches:	38
Query Match:	9.10%	Indels:	39
DB:	12	Gaps:	7

US-10-691-383-2\_COPY\_435\_632 (1-198) x BI761471 (1-794)

Qy	15	AlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArg	34
Db	315	GCCCGCTTCAGTCCGCCGCCGCTCTGTATC-----CTCAGCAAC	353
Qy	35	ProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGly-----AspLeuAspAla	52
Db	354	CCCGAG-----GGCACCTGTGAGGACACTCGGCTGGCTCTGGGCTCTCATGCA	404
Qy	53	AspPheAspIleSerLeuLeuGluAsnAspGlu-----Leu	64
Db	405	GACCACCTCTCGCCCTGCTCGAGAGCCCAAGGCCCTGACCCCGGCGCTGAGTGGCTG	464
Qy	65	LeuLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsn-----GluVal	81
Db	465	CTCAGAGGATGACAGGCCCGCGGCTGCGGCCAGAACCCCAAGACGGCTCGGTAGATATC	524

```

Qy 82 ThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSer 101
Db 525 CCTCAGCTGCTGGAGGAGGGGGTGGGGCGGGGGCTCCG----- 563
Qy 102 GlyHisAlaThrGlnAanGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAsp 121
Db 564 -----GGCAGTGTCTGGCGGGCTCTCTGCTGCCCTGCTGGACCATGTC 605
Qy 122 ArgGlyGlyGluCysPhe-----ProAsnProValPhe 132
Db 606 AGGAGCGGGTCTTGCTTCCACGCGCTTCCGAGCGCTCAGTAC 647

RESULT 5
AW661332
LOCUS      833004D10.x1 C. reinhardtii CC-125 -S, Lambda Zap II Chlamydomonas
DEFINITION reinhardtii cDNA, mRNA sequence.
ACCESSION AW661332
VERSION    AW661332.1 GI:7427164
KEYWORDS   EST.
SOURCE     Chlamydomonas reinhardtii
ORGANISM   Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE  1 (bases 1 to 558)
AUTHORS    Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
            Silflow,C., Stern,D. and Surzycki,R.
TITLE      Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants
JOURNAL    Unpublished (2000)
COMMENT     Contact: Elizabeth H. Harris
            DCMB Box 91000
            Durham, NC 27708-1000, USA
            Tel: 919 613 8164
            Fax: 919 613 8177
            Email: chlamy@duke.edu.

FEATURES             Location/Qualifiers
     source           1..558
                     /organism="Chlamydomonas reinhardtii"
                     /mol_type="mRNA"
                     /strain="CC-125 wild type mt+ 137c"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-125 -S, Lambda Zap II"
                     /notes="vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
                     XhoI; Strain CC-125 was grown to mid-log phase in TAP
                     medium, washed twice in TAP-S (sulfur-deficient TAP
                     medium) and incubated in TAP-S. RNA was isolated after 1,
                     2 and 4 hours. An equal amount of RNA from each time point
                     was pooled and polyA mRNA purified. cDNA was synthesized
                     and directionally cloned into lambda ZAP II (Stratagene)
                     in the EcoRI (5') and XhoI (3') sites. pBluescript II SK-
                     plasmids were excised from the lambda ZAP clones by
                     superinfection with ExAssist (Stratagene) phage."

ORIGIN
Alignment Scores:
Pred. No.:      2.01      Length:      558
Score:          93.50      Matches:      39
Percent Similarity: 30.53%      Conservative: 19
Best Local Similarity: 20.53%      Mismatches: 61
Query Match:     9.00%      Indels:      71
DB:              10        Gaps:         6

US-10-691-383-2_COPY_435_632 (1-198) x AW661332 (1-558)

Qy 6 SerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGln 25
Db 34 GCCCGGTCTCTCAAGCTGCTGGCTCTCTTCCTTTGACCGCGCGGCTGCTTCGT 93
Qy 26 LysTrpGlnValHisArgPheAlaArgProGluAlaLeuGlyThrLeuHisAsnThr 45

```

```

Db 94 ATCAAGTACCGGGAGCTCTTCTGGCGCCCATCACCGCCATCGCCACCATCGC--- 150
Qy 46 IleAlaGlyAspLeuAspAlaPheAspIleSerLeuLeuGluAsnAspGluLeuLeu 65
Db 151 ---GTCGGCAGCGCTGACCCACCTGG----- 174
Qy 66 LysArgValAlaGluLeuAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeu 85
Db 175 -----ACGCGCTGCTG 186
Qy 86 ProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThr 105
Db 187 -----GCCACCCCGCCCGAGTTCCTCCCTCCGCGCACCATGTC 228
Qy 106 GlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGly----- 119
Db 229 AGCTCGGGCGCTGGACTGCCATGGGGGGGGTACTTTCGGCGCCCTGGCGCCCTGAAC 288
Qy 120 -----LeuAspArgGlyGlyGluCysPheProAsnProValPhePro 133
Db 289 GTAACCACTACGGCGCCCTGGACATCTGCCCGCGCACCTTCCCT----- 333
Qy 134 SerAspGlyLeuLeuLeuIleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGlu 153
Db 349 -----GCCGTGAGTGCCTGACAGCGCGCTGTACGCTGCTGTCCACTTCAAC 396
Qy 174 PheAspGlyIleGlnGlyLeuLeuGly 183
Db 397 AGTCCAACTGACCGATTCAATCTCGGC 426

RESULT 6
BZ579009
LOCUS      1207 bp DNA linear GSS 17-DEC-2002
DEFINITION msh2_6117.x1 msh Pseudomonas aeruginosa genomic clone msh2_6117,
            genomic survey sequence.
ACCESSION BZ579009
VERSION    BZ579009.1 GI:27214070
KEYWORDS   GSS.
SOURCE     Pseudomonas aeruginosa
ORGANISM   Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE  1 (bases 1 to 1207)
AUTHORS    Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE      Whole-Genome-Sequence variation among multiple isolates of
            Pseudomonas aeruginosa library
JOURNAL    J. Bacteriol. (2002) in press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145 Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.

FEATURES             Location/Qualifiers
     source           1..1207
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="MSH"
                     /db_xref="taxon:287"
                     /clone="msh2_6117"
                     /clone_lib="msh"
                     /note="Environmental isolate. Whole genomic shotgun
                     library."

ORIGIN

```



```

Db      413  -----TATCGATATCGCGTATG 430
Qy      156  LysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyLeHisTyrArgPheAsp 175
Db      431  GTGGTGGCAGCGGTGGTGTATGAGTCGTGTGCAATGGAGCGCCATTATTTGAGTGTAT 490
Qy      176  GlyfLeGlnGlyLeuLeuGly-----GluThrIleThrValArgThrLeu 191
Db      491  GCAATCGGAGGTCTCTATTATGGGCTGCTGTTGAAGGTTTAAACAGTGTCTGCGCCCTG 544

RESULT 8
CB977242
LOCUS      818 bp      mRNA      linear      EST 01-MAY-2003
DEFINITION      CAB40003_Ia_Fa_A05 Cabernet Sauvignon Berry - CAB4 Vitis vinifera
                cDNA clone CAB40003_Ia_Fa_A05 5', mRNA sequence.
ACCESSION      CB977242
VERSION
KEYWORDS      EST.
SOURCE
ORGANISM      Vitis vinifera
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
                rosids; Vitaceae; Vitis.
REFERENCE      1 (bases 1 to 818)
AUTHORS      Goes da Silva, F., Iandolino, A., Lim, H., Baek, J., Jones, K. and
                Cook, D.
TITLE      Expressed sequence tags from Vitis vinifera 'Cabernet sauvignon'
                berries at various developmental stages
JOURNAL
COMMENT      Unpublished (2003)
                Contact: Douglas Cook, PhD
                CAES Genome Facility
                UC Davis, Plant Pathology
                One Shields Ave, Davis, CA 95616, USA
                Tel: 530 754 6561
                Fax: 530 754 6617
                Email: drcook@ucdavis.edu
                Seq primer: ACGGTACCGACATATGCC.
                Location/Qualifiers
FEATURES
    source
        1..818
            /organism="Vitis vinifera"
            /mol_type="mRNA"
            /cultivar="Cabernet Sauvignon"
            /db_xref="taxon:29760"
            /clone="CAB40003_Ia_Fa_A05"
            /sex="Hermaphrodite"
            /dev_stage="Berry on stage II, 9 mm"
            /lab_hosts="DH5alpha"
            /clone_lib="Cabernet Sauvignon Berry - CAB4"
            /notes="Organ: Berry; Vector: pDNR; Site_1: Sfil; Site_2:
            Sfil; CAB4 is a cDNA library of Vitis vinifera cv.
            'Cabernet Sauvignon'. Clone 8 berries. Sampled berries were
            collected from field-grown vines during stage II of berry
            growth (berries were green and hard) at approximately 60
            days after full bloom. The average berry size was 9
            millimeters. Sampled vines were located at the University
            of California, Davis, Experimental Vineyard. cDNAs were
            made by oligo-dT priming and directionally cloned. 5' and
            3' adaptors were used in cloning as follows:
            5'-AAGCTGATGATCAACGAGATGGCCATTACGCGCGG-3' and
            5'-ATTCTAGAGCGGAGCGGCGCATG-AT(30)NN-3'. Library was
            constructed using the Clontech Creator SMART kit and
            size-selected to contain the 0.5-3 kb size fraction."

ORIGIN
Alignment Scores:
Pred. No.:      5.66      Length:      818
Score:          92.00      Matches:      32
Percent Similarity: 45.37%      Conservative: 17
Best Local Similarity: 29.63%      Mismatches: 45
Query Match:    8.85%      Indels:      14
DB:             14      Gaps:        5

```

```

US-10-691-383-2_COPY_435_632 (1-198) x CB977242 (1-818)

Qy      38  LeuGlyGlyThrLeu-----HisAsnThrIleAlaGlyAspLeuAspAlaAspPhe 54
Db      450  CTTGGTGGATTTGTGGGGTTCTCTGATAATGATCCAGAGAGTATATTCGGTGGAT--- 506
Qy      55  AspIleSerLeuLeuLeuAsnAspGluLeuLeuLysArgValAlaGluIleAsnAlaAla 74
Db      507  GATGAATAATTACTTAAGATAGACCAATCTGGATAGGGTCTTGTCTATTTCAGCAGGC 566
Qy      75  GlnAsnProAsnAsnGluValThrTyrLeuLeu-----ProGlnAlaIleGlnValGly 92
Db      567  GTAATTGCAATAATAATAATTATTCCTATGCTATGATAATCTTTGTTCAAGTATTGTCTGTGA 626
Qy      93  SerProThrHisProSerTyrProSerGlyHisAlaThrClnAsnGlyAlaPheAlaThr 112
Db      627  TTGCTGTGTACAGAGCGCTTTCCT-----GGGGTGTCTGTGGCCT 665
Qy      113  ValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCysPheProAsnProValPhe 132
Db      666  GAGGTTGCGAGCTTGTGCGCTGCTCCGAGATGGA---TTGCTTCCAGCGCATATAATT 722
Qy      133  ProSerAspAspGlyLeuGluLeu 140
Db      723  CCTGCTGTTAATGCTATTGAATTG 746

RESULT 9
CF192410
LOCUS      537 bp      mRNA      linear      EST 08-AUG-2003
DEFINITION      14f06j2.f1 Cryptococcus neoformans strain B3501 Cryptococcus
                neoformans var. neoformans cDNA clone 14f06j2 3', mRNA sequence.
ACCESSION      CF192410
VERSION      CF192410.1 GI:33514279
KEYWORDS      EST.
SOURCE
ORGANISM      Cryptococcus neoformans var. neoformans (Filobasidiella neoformans
                var. neoformans)
                Cryptococcus neoformans var. neoformans
                Eukaryota; Fungi; Basidiomycota; Hymenomycetes;
                Heterobasidiomycetes; Tremellomycetidae; Tremellales; Tremellaceae;
                Filobasidiella.
REFERENCE      1 (bases 1 to 537)
AUTHORS      Kupfer, D.M., Drabenstot, S.D., Buchanan, K.L., Lai, H., Dyer, D.W.,
                Roe, B.A. and Murphy, J.W.
TITLE      Comparison of highly conserved intronic and exonic elements
                associated with splicing among five diverse fungal organisms
                Unpublished (2003)
JOURNAL
COMMENT      Contact: Murphy, JW
                Department of Microbiology and Immunology
                University of Oklahoma Health Sciences Center
                Oklahoma City, OK 73190, USA
                Tel: 405-271-2133 ex2133
                Email: juneann-murphy@ouhsc.edu
                This clone is available from the Fungal Genetics Stock Center,
                Phone 913-588-7044, http://www.fgsc.net/ Contact Dr. Bruce Roe
                (broe@ou.edu, www.genome.ou.edu) for sequencing questions Contact
                Dr. Juneann Murphy (juneann-murphy@ouhsc.edu) for library
                information.
                Seq primer: M13-20
                High quality sequence stop: 439.
                Location/Qualifiers
FEATURES
    source
        1..537
            /organism="Cryptococcus neoformans var. neoformans"
            /mol_type="mRNA"
            /db_xref="taxon:40410"
            /clone="14f06j2"
            /lab_hosts="E. coli strain SOLR"
            /clone_lib="Cryptococcus neoformans strain B3501"
            /note="Vector: pBluescript sk-; Site_1: EcoRI at 5' end of
            cDNA insert; Site_2: XhoI at 3' end of cDNA insert"

ORIGIN
Alignment Scores:

```

```

Pred. No.: 3.24 Length: 537
Score: 91.50 Matches: 39
Percent Similarity: 35.26% Conservative: 22
Best Local Similarity: 22.54% Mismatches: 50
Query Match: 8.81% Indels: 62
DB: 14 Gaps: 6

US-10-691-383-2_COPY_435_632 (1-198) x CF192410 (1-537)

QY 3 PheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSer 22
Db 47 TTTCTAGACACCGTTCTTCTGTTACAGTTTCGGCCACCACTGCCACTCATGGACTAGT 106
QY 23 CyTyGlnLysTrp---GlnValHisArgPhe----- 32
Db 107 CAGAACCAACGCTGGTAACAGCCCATAGATACCTCGACCATCGGGCCCGCACCTTCAG 166
QY 33 -----AlaArg-ProGluAla----- 37
Db 167 GTGTTTACGCCACCAAGCAGGCAAGGTCCAAGGACCTCTCATCGCTTTGTGAAGCA 226
QY 38 -----LeuGlyGlyThrLeuHisAsnThrIleAlaGlyAs 49
Db 227 ACAAGAAGTGAGTGAACCCGATCTTGGCGGCAAGTCCATGATACCGTAGCTGATGA 286
QY 49 pLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuLysArgValAl 69
Db 287 TCTCAGCGTTAACAACACTGATGAGGTTAATGCCGTCAGTGAGGATCCAGATCAGCGCTTGC 346
QY 69 aGluIleAsnAlaAlaGlnAsnProAsnGluAluValThrLeuLeuProGlnAlaI 89
Db 347 CGATGGAATCAAGATGATCATGGAGGAGAGAGGTAGAACAGCCCTTGAGTAGGGCGAG 406
QY 89 eGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAl 109
Db 407 CTCTGTG-----AACGGAGG 421
QY 109 aPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGly----- 123
Db 422 CTTTTCGACC-----TTTGGCGAGAAGCAGCGTCCAGAGCGCCCAAGATG 466
QY 124 -----GlyGluCysPheProAsnProValPhePro 133
Db 467 CAAGATCGGCAACTGCGAACCAAGCCCACTGTGTCCC 503

RESULT 10
CA239663/c 615 bp mRNA linear EST 25-SEP-2003
LOCUS SCBFPL5067C03.b Saccharum officinarum FL5 Saccharum officinarum
DEFINITION cDNA clone SCBFPL5067C03 3', mRNA sequence.
ACCESSION CA239663
VERSION 1
KEYWORDS EST.
SOURCE Saccharum officinarum
ORGANISM Saccharum officinarum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Saccharum.
1 (bases 1 to 615)
Vettore,A.L., da Silva,F.R., Kemper,E.B. and Arruda,P.
The libraries that made SUCST
Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
Contact: Arruda P
Centro de Biologia Molecular e Engenhariaia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parruda@unicamp.br
Clone distribution: clone distribution information can be found
through the Brazilian Clone Collection Center (BCCC) at
http://www.bcccenter.fcav.unesp.br
Plate: 067 row: C column: 03

```

Seq primer: SP6 Promoter primer.  
Location/Qualifiers  
1. .615  
/organism="Saccharum officinarum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4547"  
/clone="SCBFPL5067C03"  
/lab\_host="DH10B"  
/notes="Organ: Developed inflorescence (20cm-long) without rachis; Vector: pSport1; Site 1: SalI; Site 2: NotI; An unidirectional cDNA library generated from [Developed inflorescence (20cm-long) without rachis]. cDNA was prepared from polyA+ mRNA using SuperScript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at http://sucst.lad.ic.unicamp.br/public"

Alignment Scores:  
Pred. No.: 5.31 Length: 615  
Score: 90.50 Matches: 31  
Percent Similarity: 44.25% Conservative: 19  
Best Local Similarity: 27.43% Mismatches: 36  
Query Match: 8.71% Indels: 27  
DB: 14 Gaps: 4

US-10-691-383-2\_COPY\_435\_632 (1-198) x CA239663 (1-615)

QY 36 GluAlaLeuGlyGlyThrLeuHisAsnThrIle---AlaGlyAspLeuAspAlaAspPhe 54  
Db 571 GATGGCATTTGGGGAACTTCCAGCACTATAGGATCTGGGGTCTGGATGCACCTCCA 512  
QY 55 AspIleSerLeuLeuGluAsnAspGluLeuLeuLysArgValAlaGluIleAsnAlaAla 74  
Db 511 GAAATTTCCCTTCCCAACCTGAGGTG-----GTGGTGGCACAACAGCAATC 464  
QY 75 GlnAsnProAsnGlnValThrLeuLeuProGlnAlaIleGlnValGlySerPro 94  
Db 463 ACACAGCCTTAACAAGACATGAGTAGTGTCTCCCTCCAGCAAT----- 419  
QY 95 ThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeu 114  
Db 418 -----GAAGATGCTGCACCTACAGCAACCCCTT 392  
QY 115 LysAlaLeuIleGlyLeu-----AspArgGlyGlyGluCysPhePro 128  
Db 391 GAACATGCTGTGGCATACTGGAAGTTACACCTGCTGCTGGTGGAGGATGTTGCACCT 332  
QY 129 AsnProValPheProSerAspAspGlyLeuGluLeuIle 141  
Db 331 TCTGCTGCGAGAAATCATGAGTGTGATGATGAGAAATATA 293

RESULT 11  
CB925226 604 bp mRNA linear EST 28-APR-2003  
LOCUS ABAL\_20\_E07.g1\_A012 Abscisic acid-treated seedlings Sorghum bicolor  
DEFINITION cDNA clone ABAL\_20\_E07\_A012 5', mRNA sequence.  
ACCESSION CB925226  
VERSION 1  
KEYWORDS EST.  
SOURCE Sorghum bicolor (sorghum)  
ORGANISM Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.  
1 (bases 1 to 604)  
Cordonnier-Pratt,M.-M., Wentzel,V., Suzuki,Y., Sugano,S., Klein,R.R., Liang,C., Sun,F., Sullivan,R., Shah,M., Buchanan,C.D., Eastman,A. and Pratt,L.H.



```

Db 145 GCGGTGAAGCCCTTCTTCGCGAGAGCGGCTGGCGCTCGCGGACCGCGTGGAGAGGAGCAAG 204
Qy 23 CysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThrLeu 42
Db 205 GCCTACATGGTGTACCGCGTGGAGGACCGCGAGCGCGAG---CTCAACGGCTTCTC 261
Qy 43 -----HisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSer 57
Db 262 AAGGAGCTGGAGGCCCATCGCACACGCTGGGCGATCACAGACGTGCGAGCTGAGCTGACC 321
Qy 58 LeuLeuGluAsnAspGluLeu---LeuLysArgValAlaGluIleAsn---AlaAlaGln 75
Db 322 AGCTGAGGAGGTGTTCTCAACATCGCGCGCGCGCGGAGGTGGAGCGCGCGCGCAGC 381
Qy 76 AsnProAsnGlnValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThr 95
Db 382 AGCGGCACACCGAGGTGACGACGTGTGGACGCGCGCGCTCAAGATCCCGGTG 441
Qy 96 HisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLys 115
Db 442 GCGCGCGAGGTG---GTGGGACACCCACCGCGCGCGCTAC-----CGC 486
Qy 116 AlaLeuIleGlyLeuAspArgGlyGly-----GluCys-----PhePro 128
Db 487 GTGGATGGGACCGGACGAGCGCGCGCTGATCATCATGAGTGGAGTGGAGCTGCCG 546
Qy 129 AsnProValPheProSerAspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeu 148
Db 547 ---CCAGCTCCCGCGAGCGCGAGCTGAG-----576
Qy 149 ThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeu 168
Db 577 -----GTCGCGCGCGCGCGCGAGGCGAGGCGAGTGTATG 606
Qy 169 GlyIle 170
Db 607 GCGGTG 612

RESULT 13
LOCUS CB331374
DEFINITION 3529_136_1_G07.y_1 3529 - 2 mm ear tissue from Schmidt and Hake
labs Zea mays cDNA, mRNA sequence.
ACCESSION CB331374
VERSION CB331374.1 GI:28910308
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
TITLE 1 (bases 1 to 629)
JOURNAL clade; Panicoidae; Andropogoneae; Zea.
COMMENT Walbot,V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 3529_136_1 row: G column: 07.
FEATURES
source
location/Qualifiers
1..629
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="B73"
/db_xref="taxon:4577"
/tissue_type="ear"
/dev_stage="2 mm"

```

```

/lab_host="E. coli XL0LR"
/clone_lib="3529 - 2 mm ear tissue from Schmidt and Hake
labs"
/site="Organ: ear; Vector: PAD-GAL4-2.1; Site 1: EcoRI;
Site 2: XhoI; RNA isolated by Hake lab. 1 million pfu
amplified. Ampicillin is the selection marker."

```

## ORIGIN

```

Alignment Scores:
Pred. No.: 9.48 Length: 629
Score: 88.50 Matches: 46
Percent Similarity: 38.37% Conservative: 20
Best Local Similarity: 26.74% Mismatches: 57
Query Match: 8.52% Indels: 49
DB: 14 Gaps: 9

US-10-691-383-2_COPY_435_632 (1-198) x CB331374 (1-629)

Qy 11 LeuIleGlyAlaAlaGluLeuAlaGlnArg---AlaSerCysTyrGlnLysTrpGlnVal 29
Db 79 CTCACCGGATGCGGCTTCTGGCGGAGCGCGGCGCTGTTCTTCGACCCATCCCGGTG 138
Qy 30 HisArgPheAla---ArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGly 48
Db 139 GTTCAACTCTCTTGGCGCGCCCGCGCGCTTC-----TTGCACAGGGTTTCTGTG 189
Qy 49 AspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLysArgVal 68
Db 190 GATCGGAGTGGCGACCACTCATCGCGCTGGCCAGGACCAAGCTGGAGAGTCTATGTG 249
Qy 69 AlaGluIleAsnAlaAlaGlnAsnProAsnGluVal-----81
Db 250 GCGGACAAACAAGTCCGGTAAGAGTGTCCAGAGTGGCGGCGACAGTTCGGCATGTTC 309
Qy 82 -----Thr 82
Db 310 CTCGAGAAGAAGCAGGATAGTAGTAACGAGGATAGAGGAGGATATCTGTTGGACA 369
Qy 83 TyrLeuLeuPro-----GlnAlaIleGlnVal-----GlySerPro 94
Db 370 TTCTTCCACGAGAGATGTTGAAGCATTCAGATATTACACTACCAAGATGGTGAGAA 429
Qy 95 ThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAla-----109
Db 430 TATGAACCCCACTACGACTACTTCCATGACAAAAATAATCAGGCCCTGGGCGGCATCGC 489
Qy 110 PheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGlu---CysPhePro 128
Db 490 ATTGCCACTGTGTGTATGTACCTATCAATGTGAGAGGGTGGAGAGACCATCTTCCCC 549
Qy 129 AsnPro-----ValPheProSerAspAsp 136
Db 550 AATCGAGAGGGAAGCTATTGCAACCCAGGACGAC 585

```

## RESULT 14

```

CD223063
LOCUS CD223063
DEFINITION CCC1_25_C01_91_A007 Callus culture/cell suspension Sorghum bicolor
cDNA clone CCC1_25_C01_A007 5', mRNA sequence.
ACCESSION CD223063
VERSION CD223063.1 GI:30966497
KEYWORDS EST.
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
TITLE 1 (bases 1 to 639)
JOURNAL Cordonnier-Pratt,M.-M., Wentzel,V., Suzuki,Y., Sugano,S.,
Klein,R.R., Liang,C., Sun,F., Sullivan,R., Shah,M., Rathore,K.,
Eastman,A. and Pratt,L.H.
An EST database from Sorghum: callus culture and cell suspension
Unpublished (2003)

```





```
Db      100  CTGCGCGGCTCGCGGCGAGCGGGGGCGCTGCTTCCTTCGACCCCTCCCGCGTG 159
Qy      30  HiEArgPheAla---ArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGly 48
      :::  :::  |||||  |||  |||||
Db      160  GTCCAACTCTCGTGGCGCCCGCGGCGTTC-----CTGCACAAGGGTTTCCTGTGCG 210
Qy      49  AspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuLys---Arg 67
      |||  :::  |||  :::  |||  :::  |||  |||  |||  |||
Db      211  GATGCGGAGTGCAGAC---CACCTCATGCTGGCCCAAGGACAACTGGAGAAAGTCTATG 267
      |||||  :::  |||  :::  |||  :::  |||  |||  |||  |||
Qy      68  ValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluVal----- 81
      |||||  :::  |||  :::  |||  :::  |||  |||  |||  |||
Db      268  GTAGCGGACAAACGAGTCCGGTAAGAGCGTCCAGAGCGAGGTGCGCACCGATGCCGCGCATG 327
      |||||  :::  |||  :::  |||  :::  |||  |||  |||  |||
Qy      81  ----- 81

Db      328  TTCCTGGAGAAGACGAGGATAGTAGTAGTAAGGGGAATAGAGGAGGATAGCTGCTTGG 387
Qy      82  ThrTyrLeuLeuPro-----GlnAlaIleGlnVal-----GlySer 93
      |||  :::  |||  :::  |||  :::  |||  |||  |||  |||
Db      388  ACATTTCTCCACCAGAGAATGGTGAATCCATTCCAGATATTACACTACCAGAAATGGTGAG 447
      |||  :::  |||  :::  |||  :::  |||  |||  |||  |||
Qy      94  ProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAla----- 109
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      448  AAATACGAACCCCACTATGACTTCCATGACAAAAAATAATCAAGCCCTGGGTGGCCAT 507
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy      110  ---PheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGlu---CysPhe 127
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      508  CGCATTCCTGCTGCTCATGTACTTATCTAATGTTGAGAGGGTGGAGAGACCATCTTC 567
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy      128  ProAsnPro-----ValPheProSerAspAsp 136
      |||||  :::  |||  |||||
Db      568  CCCAATGCAGAGGGAAGCTATTACAGCCCAAGGACGAC 606
```

Search completed: September 19, 2004, 09:22:46  
Job time : 1601.65 secs

**This Page Blank (usp10)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2004, 01:14:52 ; Search time 23.9771 Seconds  
(without alignments)  
1468.042 Million cell updates/sec

Title: US-10-691-383-2

Perfect score: 3528

Sequence: 1 MLCHAADTRGSPMPDTGVL.....SIDGDMCSGLVYTGVDQCA 676

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query Match %	Score	Length	ID	Description
1	2514	71.3	557	1	PRXV_ASCNO
2	114.5	3.2	690	1	NRF_HUMAN
3	113.5	3.2	912	1	CAD2_CHICK
4	110.5	3.1	439	1	COMP_BOVIN
5	107	3.0	1819	1	GCP6_HUMAN
6	106.5	3.0	947	1	TOPI_MYCLE
7	106.5	3.0	1016	1	DP01_RHILE
8	106	3.0	303	1	LPXC_PSESM
9	106	3.0	426	1	HISX_METH
10	106	3.0	3670	1	CSM3_HUMAN
11	104.5	3.0	554	1	SECD_RHOCA
12	104.5	3.0	615	1	MUTA_MYCTU
13	104.5	3.0	1305	1	FTSK_YERPE
14	103.5	2.9	755	1	COMP_MOUSE
15	103.5	2.9	755	1	COMP_RAT
16	103.5	2.9	757	1	COMP_HUMAN
17	102.5	2.9	469	1	GLNA_VIBCH
18	102	2.9	737	1	PLO2_HUMAN
19	101.5	2.9	567	1	SYFB_HALN1
20	101.5	2.9	644	1	FTSH_ECO57
21	101.5	2.9	644	1	FTSH_ECOLI
22	101.5	2.9	906	1	CAD2_MOUSE
23	101.5	2.9	1036	1	HP12_DEIRA
24	101	2.9	741	1	HLVA_VIBCH
25	101	2.9	877	1	CAD2_BOVIN
26	101	2.9	1744	1	TENS_CHICK
27	100	2.8	566	1	ATFI_SCHPO
28	100	2.8	625	1	DXS_HAEIN
29	100	2.8	877	1	SECA_GUTH
30	100	2.8	948	1	HP11_DEIRA
31	100	2.8	2124	1	PGCA_RAT
32	100	2.8	2319	1	NTC3_RAT
33	99.5	2.8	791	1	TEX_BORPE

## RESULT 1

ID	PRXV_ASCNO	STANDARD;	PRT;	557 AA.
AC	P81701;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Vanadium haloperoxidase (EC 1.11.1.-) (V-BPO).			
OS	Ascopyllium nodosum (knotted wrack) (Brown seaweed).			
OC	Eukaryota; stramenopiles; Phaeophyceae; Fucales; Fucaea;			
OC	Ascopyllium.			
OX	NCBI_TaxID=52969;			
RN	[1]			
RP	SEQUENCE, X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS), AND FUNCTION.			
RX	MEDLINE=20013071; PubMed=10543953;			
RA	Weyand M., Hecht H.-J., Kiess M., Liaud M.-F., Vilter H.,			
RA	Schomburg D.;			
RT	"X-ray structure determination of a vanadium-dependent			
RT	haloperoxidase from Ascopyllium nodosum at 2.0-A resolution."			
RL	J. Mol. Biol. 293:595-611(1999).			
RN	[2]			
RP	SEQUENCE OF 320-556 FROM N.A., SEQUENCE OF 326-341; 383-426; 471-479			
RP	AND 481-556, AND FUNCTION.			
RX	MEDLINE=96081028; PubMed=8564812;			
RA	Vilter H.;			
RT	"Vanadium-dependent haloperoxidases."			
RL	(in) Sigel H., Sigel A. (eds.);			
RL	Metal ions in biological system-vanadium and its role in life,			
RL	pp.31:325-362, Marcel Dekker, New York (1995).			
CC	-1- CATALYTIC ACTIVITY: Halide + H(2)O(2) + H(+) = HOHal + H(2)O.			
CC	-1- COFACTOR: Vanadium.			
CC	-1- SUBUNIT: Homodimer; disulfide-linked.			
CC	-1- SIMILARITY: TO OTHER BACTERIAL NON-HEME BROMO- AND CHLORO-			
CC	PEROXIDASES.			
DR	PDB; 1Q19; 10-JUN-00.			
DR	InterPro; IPR008934; AcPase Vanperase.			
DR	InterPro; IPR000326; PA_PTPase.			
KW	Oxidoreductase; Peroxidase; Vanadium; 3D-structure;			
KW	Pyridinolone carboxylic acid.			
FT	MOD RES 1 1			
FT	DISULFID 3 3			
FT	DISULFID 41 41			
FT	DISULFID 77 86			
FT	DISULFID 441 462			
FT	DISULFID 544 555			
FT	ACT SITE 411 411			
FT	ACT SITE 418 418			
FT	METAL 486 486			
FT	CONFLICT 321 321			
FT	CONFLICT 341 341			
FT	CONFLICT 403 404			
FT	CONFLICT 407 408			
FT	CONFLICT 409 409			
FT	CONFLICT 441 444			
FT	CONFLICT 470 470			
FT	HELIX 15 37			

P78871 schizosacch  
Q53114 mycobacteri  
Q92fi8 rhodobacter  
Q8xg72 salmonella  
Q921v3 rattus norv  
P50851 homo sapien  
P27050 bacillus ci  
P35442 homo sapien  
P21241 brassica na  
Q89af2 buchnera ap  
Q50596 mycobacteri  
Q03043 drosophila

## ALIGNMENTS

50 52  
53 53  
54 53  
55 57  
56 59  
57 60  
58 60  
59 62  
60 63  
61 66  
62 66  
63 80  
64 87  
65 88  
66 88  
67 98  
68 100  
69 102  
70 103  
71 108  
72 108  
73 118  
74 119  
75 129  
76 130  
77 132  
78 132  
79 146  
80 147  
81 149  
82 152  
83 154  
84 155  
85 157  
86 170  
87 170  
88 172  
89 173  
90 174  
91 176  
92 178  
93 179  
94 180  
95 184  
96 185  
97 190  
98 195  
99 197  
100 197  
101 200  
102 203  
103 203  
104 210  
105 213  
106 216  
107 218  
108 219  
109 220  
110 221  
111 223  
112 227  
113 227  
114 230  
115 230  
116 236  
117 236  
118 240  
119 247  
120 249  
121 248  
122 258  
123 258  
124 263  
125 263  
126 268  
127 268  
128 277  
129 282  
130 294  
131 295  
132 299  
133 300  
134 304  
135 308  
136 312  
137 313  
138 315  
139 315  
140 317  
141 318  
142 318  
143 320  
144 329  
145 330  
146 333  
147 334  
148 342  
149 342  
150 343  
151 347  
152 350  
153 361  
154 362  
155 371  
156 374  
157 375  
158 375  
159 377  
160 377  
161 390  
162 393  
163 393  
164 402  
165 402  
166 406  
167 406  
168 417  
169 433  
170 433  
171 435  
172 438  
173 439  
174 439  
175 441  
176 441  
177 446  
178 448  
179 450  
180 451  
181 455  
182 457  
183 464  
184 464  
185 464  
186 464  
187 464  
188 464  
189 484  
190 484  
191 488  
192 488  
193 512  
194 519  
195 522  
196 524  
197 525

FT STRAND 528 531  
FT TURN 533 534  
FT STRAND 537 539  
FT TURN 540 541  
FT STRAND 542 543  
FT HELIX 552 554  
SQ SEQUENCE 557 AA; 60343 MW; E3D8557AB92B16F4 CRC64;  
  
Query Match 71.3%; Score 2514; DB 1; Length 557;  
Best Local Similarity 86.5%; Pred. No. 1.8e-177;  
Matches 480; Conservative 31; Mismatches 40; Indels 4; Gaps 1;  
  
QY 124 QTCTNSDDALDPTAPNRDNVAFASRRDAAARRRDGTGTVCQITNGTDTLATMFKSLPH 183  
DB 1 QTCTSDDDADDPTPPNERDDEAFASRAAKRELGTTGTCQINNGETDLAAKFKSLPH 60  
  
QY 184 DELQVTTADDFATLEDCILNGDFSI CEDVPA----GDPAGRLVNPATAFAIDISGPAPSA 239  
DB 61 DDLGQVDADAFALDCEILNGDLSCIEDVPVGNSEGDVPVGRVNPATAFAIDISGPAPSA 120  
  
QY 240 TTTPPVPTLSSPELAQAQLAEVYMWALARDVPFMOYGTDEITTTAAANLAGMGFPNLDV 299  
DB 121 TTTPPVPTLSPPELAQAQLAEVYMWALARDVPFMOYGTDDITVTAANLAGMGFPNLDV 180  
  
QY 300 SIGSDGTVDPSQLFRATFVGTVETGPFVSVQLLVNSFTIDAITYEPKQETAPDLNVMVF 359  
DB 181 SIGSDGTVDPSQLFRATFVGTVETGPFVSVQLLVNSFTIDSITYEPKQETAPDLNVMVF 240  
  
QY 360 DEWLNTQNGGPPAGPELDBELFIRNARDLARVSVDNINTAYRGSLLLELGAFSRP 419  
DB 241 DEWLNTQNGGPPAGPELDDDELFRVNRNARDLARVTDTNINTAYRGALLGLDAFNRA 300  
  
QY 420 GINGPFTDSROAGFVNFSGTSHYFRILGAEALQASRCYKQWQVHRFARPEALGGTLHNT 479  
DB 301 GVGPFIDIDRQAGFVNFSGTSHYFRILGAEALQASRCYKQWQVHRFARPEALGGTLHNT 360  
  
QY 480 IAGDLDAFDSLLENDELKRVAEINAAQNPNEVTLTPQAIQVGSPTSPSPSGHAT 539  
DB 361 IKGELNADFDLSLENAELLKRVAAINAAQNPNEVTLTPQAIQVGSPTSPSPSGHAT 420  
  
QY 540 QNGAFATVTKALIGDRGCECFNPVPSDDGLELINFEGACLTVEGEINKLAVNAFGR 599  
DB 421 QNGAFATVTKALIGDRGCECFNPVPSDDGLELINFEGACLTVEGEINKLAVNAFGR 480  
  
QY 600 QMLGIHYRFDGIQGLLGETITVTTHOELMTAEERATFEFRFLTGTGVIKLFDGTFSD 659  
DB 481 QMLGIHYRFDGIQGLLGETITVTTHOELMTAEERATFEFRFLTGTGVIKLFDGTFSD 540  
  
QY 660 GDMCSGLVYTGVDAC 674  
DB 541 GFKCPGLVYTGVCNC 555

RESULT 2

NRF\_HUMAN  
ID NRF\_HUMAN STANDARD; PRT; 690 AA.  
AC O15226; Q9UJ91;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE NF-kappa-B-repressing factor (NFKB-repressing factor) (Transcription factor NRF) (ITBA4 protein).  
DE factor NRF (ITBA4 protein).  
GN NRF OR ITBA4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N. A.  
RC TISSUE=Cervical carcinoma;  
RX MEDLINE=20031652; PubMed=10562553;  
RA Nourbakhsh M., Hauser H.;  
RT "Constitutive silencing of IrN-beta promoter is mediated by NRF (NF-

632  
-435 21 = 0.106  
197 198  
89.40%



SEQUENCE FROM N.A.  
 MEDLINE=88153917; PubMed=2831236;  
 RA Hatta K., Nose A., Nagafuchi A., Takeichi M.;  
 RT "Cloning and expression of cDNA encoding a neural calcium-dependent  
 cell adhesion molecule: its identity in the cadherin gene family."  
 RL J. Cell Biol. 106:873-881(1988).  
 RN [2]  
 RP SEQUENCE OF 1-25 FROM N.A.  
 RC STRAIN=Cornish white rock Cockerel;  
 RX MEDLINE=97354288; PubMed=9210582;  
 RA Li B., Paradies N.E., Brackenbury R.W.;  
 RT "Isolation and characterization of the promoter region of the chicken  
 N-cadherin gene."  
 RL Gene 191:7-13(1997).  
 CC -1- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.  
 CC They preferentially interact with themselves in a homophilic  
 CC manner in connecting cells; cadherins may thus contribute to the  
 CC sorting of heterogeneous cell types. N-cadherin may be involved in  
 CC neuronal recognition mechanism.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: Contains 5 cadherin domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X07277; CAA30258.1; -;  
 DR EMBL; U15563; AAB62980.1; -;  
 DR PIR; A29364; IJCHCN.  
 DR HSPSP; P15116; INCJ.  
 DR InterPro; IPR002126; Cadherin.  
 DR InterPro; IPR002333; Cadherin\_C term.  
 DR Pfam; PF00028; cadherin; 5.  
 DR Pfam; PF01049; Cadherin\_C term; 1.  
 DR PRINTS; PR00205; CADHERIN.  
 DR SMART; SM00112; CA; 5.  
 DR PROSITE; PS00232; CADHERIN 1; 3.  
 DR PROSITE; PS0268; CADHERIN 2; 5.  
 DR Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;  
 KW Signal.  
 FT SIGNAL. 1 28  
 FT PROPEP 29 164  
 FT CHAIN 165 912  
 FT DOMAIN 165 729  
 FT TRANSMEM 730 752  
 FT POTENTIAL.  
 FT DOMAIN 753 912  
 FT CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 165 272  
 FT CADHERIN 1.  
 FT DOMAIN 273 387  
 FT CADHERIN 2.  
 FT DOMAIN 388 502  
 FT CADHERIN 3.  
 FT DOMAIN 503 609  
 FT CADHERIN 4.  
 FT DOMAIN 610 720  
 FT CADHERIN 5.  
 FT DOMAIN 869 884  
 FT SER-RICH.  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 278 278  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 330 330  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 407 407  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 578 578  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 628 628  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 657 657  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 21 21  
 FT A -> G (IN REF. 2).  
 SQ SEQUENCE 912 AA; 100464 MW; 9BAC9DC1FFC489 CRC64;  
 Query Match 3.28; Score 113.5; DB 1; Length 912;  
 Best Local Similarity 21.28; Pred. No. 2.4;  
 Matches 113; Conservative 65; Mismatches 195; Indels 160; Gaps 28;  
 QY 175 TMFKSLPH-----DELQVTTADDPFAILEDCLNGDPSICEDVPAGDPAAGLVNP----- 224  
 Db 409 TVTDKQDPHTPAMNARYQMTGGD-----PTGQFTILTDPNSNDGLVTWVKPDPFTN 460

QY 225 -----TAA-----FAIDISGPAFSATTIP-----PVPTLSSEPAAQAE 260  
 Db 461 RMFVLTVAAENQVPLAKGQHPPQSTATVSIITVIDVNESYFVPNPKLVQEESGLAGSM 520  
 QY 261 YMWALARDVP-FWQYGTDEITTTAAANLAGMGFFNLDAVS---IGSDGTVDPPFS---Q 312  
 Db 521 LTTFTTARDPRVMQ-----TSLSYSKLSDPANLWKIDPVGQITTTAVLDRESIYVQNN 575  
 QY 313 LFRATFVGVEG-PFVS-----QLLVNSFTIDAITVEPKQ-----ETFPADLVNVMVDEW 362  
 Db 576 MYNATFLASDNGIPMSGTGLQIYLLDINDNAPQVNPKEATTCETLQPNAINITAVDPD 635  
 QY 363 LNTQNGCP-----PAGPEELDEELRIRNARDLARVS-----FVDNINT 401  
 Db 636 ID-PNAGPAPAFELPDSPPSIKRNTVIRISGDHAQLSLRFLFLEAGIYDVPITVDSGNP 694  
 QY 402 EAYRGLLLELGAFRSPGNGINGPFIQSDRQAGFVNGFTSHYFR-LIGAABLAORASYQK 460  
 Db 695 HASSTSVLKVK-----CQCIDNGDCTDVRIVG-AGLGTGAIILLCIILLILVLMFVV 750  
 QY 461 WQVHREARPEAL-----CGTLHNTIAGLDADDFDISLLENDELLK 500  
 Db 751 WKRRDKERQAKOLLIDPDVDRDNILKYDEEG-----GEEDQYDLSLQOQPTVE 803  
 QY 501 RVAEINAAQNPNNEVYLLPQAIQ-VG-----SPTH--PSYP-----SGHATONGAFATV 547  
 Db 804 -----PDAIKPVGIRLDERPIHAEPQYFVRSAAHPGIDIGDF--I 842  
 QY 548 LKALIGLDRGEGCFNPVFPSPDGLLEINPEGACLYVEGINKLVANVAFGRQ 600  
 Db 843 NEGLKAADN-----DPTAPPYDLSLLVDFYEGSGST-AGSLSLNSSSGGEQ 888  
 RESULT 4  
 COMP BOVIN STANDARD; PRT; 439 AA.  
 AC P35445;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Cartilage oligomeric matrix protein (COMP) (Fragment).  
 GN COMP.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Cartilage;  
 RX MEDLINE=93054522; PubMed=1429587;  
 RA Oldberg A., Antonsson P., Lindblom K., Heinegaard D.;  
 RT "COMP (cartilage oligomeric matrix protein) is structurally related  
 to the thrombospondins."  
 RL J. Biol. Chem. 267:22346-22350(1992).  
 CC -1- SUBUNIT: Pentamer; disulfide-linked.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: Belongs to the thrombospondin family.  
 CC -1- SIMILARITY: Contains 7 TSP type-3 domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X74326; CAA52374.1; -;  
 DR PIR; B44315; B44315.  
 DR InterPro; IPR003367; tsp\_3.  
 DR InterPro; IPR008859; TSPC.  
 DR Pfam; PF02412; tsp\_3; 11.

DR Pfam; PF05735; TSPC; 1.  
 KW Glycoprotein; Cell adhesion; Calcium-binding; Repeat.  
 FT NON TSP 1  
 FT DOMAIN <1 14 TSP TYPE-3 1.  
 FT DOMAIN 15 37 TSP TYPE-3 2.  
 FT DOMAIN 38 73 TSP TYPE-3 3.  
 FT DOMAIN 74 96 TSP TYPE-3 4.  
 FT DOMAIN 97 134 TSP TYPE-3 5.  
 FT DOMAIN 135 170 TSP TYPE-3 6.  
 FT DOMAIN 171 206 TSP TYPE-3 7.  
 FT DOMAIN 207 439 C-TERMINAL.  
 FT CARBOHYD 424 424 N-LINKED (GLCNAC... (POTENTIAL).  
 SQ SEQUENCE 439 AA; 49962 MW; 9CE8BCDD0D98070 CRC64;  
 Query Match 3.1%; Score 110.5; DB 1; Length 439;  
 Best Local Similarity 24.6%; Pred. No. 1.4;  
 Matches 52; Conservative 19; Mismatches 89; Indels 51; Gaps 10;  
 QY 112 BEHKTGEGHLCQCTNSDADLPTAPNRDNNVAFSPRRDAARRERDGTGTCQITNGET 171  
 DB 139 QDSHDHGGQ---DACDDDD--NDGVPSDRNCRVLPNPGQDMDRDGVGDACQ---GDF 190  
 QY 172 DLATMFHKLPHDELGVGTADDFAILEDCLN--GDPSI-----CEDVPA--GDP 217  
 DB 191 DAKVDVKIDVCENAEVITLDFRAFQTVVLDPEGDAQIDPNVVLNQGHEIVQTMNSDP 250  
 QY 218 AGRLVNPTAAFAIDISGPAFSATTIPVPTLSSPELAA-----QLAEL 260  
 DB 251 -GLCVGYTAFNGVDFEGPFH-----VNTATDDYAGTFPGYHSSSFVVMKQMEQT 302  
 QY 261 YWALARDVPMQVGTDEITTTAAANLACMG 291  
 DB 303 YWQA---NPFRAVAPGPIQLKAVKSTGPG 329  
 RESULT 5  
 GCP6 HUMAN STANDARD; PRT; 1819 AA.  
 AC Q96RT7; Q9BY91; Q9UGX3; Q9UGX4;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Gamma-tubulin complex component 6 (GCP-6).  
 GN TUBGCP6 OR GCP6 OR KIAA1669.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND CHARACTERIZATION.  
 RX MEDLINE=21551508; PubMed=11694571;  
 RA Murphy S.M., Preble A.M., Patel U.K., O'Connell K.L., Dias D.P.,  
 RA Moritz M., Agard D., Stules J.T., Stearns T.,  
 RT "GCP5 and GCP6: two new members of the human gamma-tubulin complex.";  
 RL Mol. Biol. Cell 12:3340-3352(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20057165; PubMed=10591208;  
 RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,  
 RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,  
 RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,  
 RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,  
 RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,  
 RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,  
 RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,  
 RA Dhani P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,  
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,  
 RA Gilbert J.G.R., Goward M.E., Graham D.V., Griffiths M.N.D., Hall C.,  
 RA Hall R.E., Hall-Tamlyn G., Heathcott R.W., Ho S., Holmes S.,  
 RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,  
 RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,  
 RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,  
 RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,

RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,  
 RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,  
 RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,  
 RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,  
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,  
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,  
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,  
 RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,  
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuayama S.,  
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,  
 RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,  
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,  
 RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,  
 RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,  
 RA Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,  
 RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,  
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,  
 RA Hinds K., Kemp K., Latreille P., Layman D., Ozeraky P., Rohlfing T.,  
 RA Scheet P., Walter C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,  
 RA Korf I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,  
 RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,  
 RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,  
 RA Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumaneki J.P.,  
 RA Peyrard M., Kedra D., Serousi E., Fransson I., Tapia I., Bruder C.E.,  
 RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,  
 RA Khan A.S., Lane L., Tiliahun Y., Wright H.;  
 RA "The DNA sequence of human chromosome 22.";  
 RL Nature 402:489-495(1999).  
 RN [3]  
 RP SEQUENCE OF 33-1819 FROM N.A. (ISOFORM 2).  
 RC TISSUE=Brain;  
 RX MEDLINE=21156230; PubMed=11258795;  
 RA Hirobawa M., Nagase T., Murahashi Y., Kikuno R., Ohara O.;  
 RT "Identification of novel transcribed sequences on human chromosome 22  
 by expressed sequence tag mapping.";  
 RL DNA Res. 8:1-9(2001).  
 CC -!- FUNCTION: Gamma-tubulin complex is necessary for microtubule  
 nucleation at the centrosome.  
 CC -!- SUBUNIT: Gamma-tubulin complex is composed of gamma-tubulin, GCP2,  
 GCP3, GCP4, GCP5 and GCP6.  
 CC -!- SUBCELLULAR LOCATION: Centrosome.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=Q96RT7-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q96RT7-2; Sequence=VSP\_001624;  
 CC Note=No experimental confirmation available;  
 CC -!- SIMILARITY: Belongs to the GCP family.  
 CC -!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts  
 in positions 1371 and 1758.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC -----  
 CC EMBL; AF272887; AK82968.1; -  
 CC EMBL; AL022328; CAB63046.1; ALT SEQ.  
 CC EMBL; AL022328; CAB63047.1; ALT SEQ.  
 CC EMBL; AB051456; BAB33339.1; ALT\_FRAME.  
 CC Genew; HGNC:18127; TUBGCP6.  
 CC GO; GO:0008274; C:gamma-tubulin ring complex; IDA.  
 CC GO; GO:0008017; F:microtubule binding; IDA.  
 CC GO; GO:0007020; P:microtubule nucleation; IDA.  
 CC InterPro; IPR007259; SPC97\_Spc98.  
 CC Pfam; PF04130; SPC97\_Spc98; 1.  
 CC Microtubule; Repeat; Alternative splicing.  
 FT DOMAIN 1027 1269 9 x 27 AA TANDEM REPEATS.  
 FT REPEAT 1027 1053 1.





```
Db 679 RANLNGTLTDPDELTVLVAELFATPHGRVLGVD-----PETGHEIVAKDGRY--- 726
Qy 387 ARDLARVSFVDNINTEAYRGSLLLELGAFSRPGINGPFIDS-----DROAGFVNFCTSH 441
Db 727 -----GYYTEVLPHKDDYGAADQTK- 749
Qy 442 YFRLIGAEALQARASCYQKQVHRFARPEALGGTLHNTIAGDLDADPDISLLENDELLKR 501
Db 750 -----KTKGRASASQG-----PKPR-----TGSLLRSMDLQTITLEDALKL 787
Qy 502 VAEINAAQNPNVETLLPQAIQVGSPTSPSGHATONGAFATVLKALIGLDRGCECF 561
Db 788 L-----LLPRVVG-----DPASGEITQNGRYGPLYK-----RGKDSR 822
Qy 562 PNPVFPSPDDGLEINFEACILITY 584
Db 823 P---LTTEDQMFITLDEALKIY 842

RESULT 7
ID DPOL RHILE STANDARD; PRF; 1016 AA.
AC Q9S1G2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 18-OCT-2001 (Rel. 40, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA polymerase I (EC 2.7.7.7) (POL I).
GN POLA.
OS Rhizobium leguminosarum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCBI_TaxID=384;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=9272512; PubMed=10341077;
RA Huang Y.-P., Downie J.A., Ito J.;
RT "Primary structure of the DNA polymerase I gene of an alpha-
RT Proteobacterium, Rhizobium leguminosarum, and comparison with other
RT family A DNA polymerases."
RL Curr. Microbiol. 38:355-359(1999).
CC -1- FUNCTION: In addition to polymerase activity, this DNA polymerase
CC exhibits 3' to 5' and 5' to 3' exonuclease activity (By
CC similarity).
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA] (N).
CC -1- SUBUNIT: Single-chain monomer with multiple functions.
CC -1- SIMILARITY: Belongs to the DNA polymerase type-A family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC -----
CC EMBL; U86403; RAD4559.1; --
CC HSSP; P00582; IKFS.
DR InterPro; IPR002562; 3_5_exonuclease.
DR InterPro; IPR008918; 5_3_exo_C.
DR InterPro; IPR002421; 5_3_exonuclease.
DR InterPro; IPR001098; DNA_pol.
DR InterPro; IPR002298; DNA_poli.
DR InterPro; IPR000513; Exo_N_I.
DR InterPro; IPR006055; Exonuclease.
DR InterPro; IPR003583; HHH_1.
DR Pfam; PF01612; 3_5_exonuclease; 1.
DR Pfam; PF02739; 5_3_exonuc N; 1.
DR Pfam; PF01367; 5_3_exonuclease; 1.
DR Pfam; PF00476; DNA_pol_A; 1.
DR PRINTS; PR00868; DNAPOLI.
DR SMART; SM00474; 3SEXOC; 1.
DR SMART; SM00475; 53EXOC; 1.
```

```
DR SMART; SM00479; EXOIII; 1.
DR SMART; SM00278; HHH1; 1.
DR SMART; SM00279; HHH2; 1.
DR SMART; SM00482; POLAC; 1.
DR TIGRfam; TIGR00593; Pola; 1.
DR PROSITE; PS00447; DNA_POLYMERASE_A; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication; DNA repair;
KW Hydrolyase; Exonuclease; DNA-binding.
FT DOMAIN 1 308 5'-3' EXONUCLEASE (BY SIMILARITY).
FT DOMAIN 394 630 3'-5' EXONUCLEASE (BY SIMILARITY).
FT DOMAIN 768 1016 POLYMERASE (BY SIMILARITY).
SQ SEQUENCE 1016 AA; 111491 MW; 35738F85B8C70514 CRC64;

Query Match 3.0%; Score 106.5; DB 1; Length 1016;
Best Local Similarity 18.8%; Pred. No. 9.1;
Matches 110; Conservative 86; Mismatches 215; Indels 174; Gaps 26;

Qy 177 FHKSPLHDELQVTAD-----DFAILEDICILN-----GDFSIDVDPAGDPAGRLVNPNTAFA 229
Db 295 FLXTMEFTTLTRRVAEACDCDAIIPAIVRIEMGETARGPDLDAEP-----EPVAGGI 349
Qy 230 IDISGPAFSATTIPP-----VPTLSSP-ELAAQLAELY----- 261
Db 350 PEVSG---ESVMPPRAKAKSAVEGAFSPADLAKARAEAFATLPFDHSAYTIRDLVTLD 406
Qy 262 -MMALARDVPFQYGTDEITTTAAANLAGMGGFNNLDVAGISGSDGTVDYDFPSQLFRATFVG 320
Db 407 RWIADARATGLVAPDTE--TTSLDAMQALVGF-----SLAI-ADNTADPTGTKIRAAVVP 459
Qy 321 VETGPFVSQLLVNSFTIDAITVEPKQETPAPDLNVMVDFEWNLIQNGGPPAGPEEDEE 380
Db 460 LVHKNVGVDLLGGGLADNQI---PMRDAL-PRLKALLEDESVLKVA-----QN 503
Qy 381 LRF-----INARDLARVSFVDNINTEAY-----RGSLLLELGAFSR 418
Db 504 LKDYDLLLKKYGIETRSFDDTMLSIVLDAGTAGMDPLSEKFLGHTPTPYKDVAGSGK 563
Qy 419 PGINGPPIIDSDRQAGFVNFCTSHYFRL-----IGAAELAAQRASCYQKQW---VHRFR 468
Db 564 ANVTFDLVIDIRATHYAAEDADVTLRLWLKPLRAAAGL---TSVYERLERPLLPVLAR 620
Qy 469 PEALGTLHNTIAGDLDADPDISLLENDELLKRVAEINAAQNPNVETLLPQAIQVGP 528
Db 621 MEARGITVDRQILSR-----SGELAAQGAARL-----EDEIYVLAGERFNIGSP 664
Qy 529 TH-----PSYPSGHATONGAFATVLKALIGLDRGG-----E 559
Db 665 KQLGDILFGKMGSGSKTKTGWSTSAQVLEDLAAAGFELPKIVDWRQVTKLKSTYTD 724
Qy 560 CFNPVPFPSPDDGLE-----LINFEAGACTYEGEINKLVNNAFGRQMLGHYRFDGIQGL 614
Db 725 ALPGYVHPETKRVHTSVSLASTTTGRLSSSEPNLQNPVTAEGRKI----- 771
Qy 615 LLETTTTRTLHQLMTFAEATPEFRLFT-----GEVILKFDQG 654
Db 772 ---RTAFISTPGHKLIS-ADYSQIELRVLVAHVAEIPOLTKAFEDG 812

RESULT 8
LPXC_PSESIM STANDARD; PRF; 303 AA.
AC Q87WZ1;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase
DE (EC 3.5.1.-) (UDP-3-O-acetyl-GlcNAc deacetylase).
GN LPXC OR PSPTO4402.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
```





```
FT DOMAIN 2595' SUSHI 16. 2652
FT DOMAIN 2657 SUSHI 17. 2717
FT DOMAIN 2722 SUSHI 18. 2775
FT DOMAIN 2780 SUSHI 19. 2833
FT DOMAIN 2838 SUSHI 20. 2891
FT DOMAIN 2896 SUSHI 21. 2953
FT DOMAIN 2958 SUSHI 22. 3011
FT DOMAIN 3019 SUSHI 23. 3072
FT DOMAIN 3077 SUSHI 24. 3131
FT DOMAIN 3136 SUSHI 25. 3191
FT DOMAIN 3196 SUSHI 26. 3249
FT DOMAIN 3254 SUSHI 27. 3307
FT DOMAIN 3315 SUSHI 28. 3369
FT DOMAIN 3374 SUSHI 29. 3429
FT CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 36 36 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 324 324 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 687 687 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 786 786 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 929 929 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1055 1055 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1134 1134 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1243 1243 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1499 1499 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1554 1554 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1672 1672 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1744 1744 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1892 1892 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1982 1982 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2118 2118 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2249 2249 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2254 2254 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2287 2287 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2458 2458 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2500 2500 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2674 2674 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2705 2705 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2825 2825 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2895 2895 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2915 2915 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3062 3062 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3121 3121 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3130 3130 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3157 3157 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3171 3171 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3181 3181 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3239 3239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3327 3327 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3485 3485 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3492 3492 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3575 3575 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3581 3581 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3590 3590 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC MGKSGFTFWNLVLLTVSCVK -> MWSWFLCWKPQLDR
FT 1 22 QTAS (in isoform 2).
FT 307 Missing (in isoform 3).
FT 410 /FTid=VSP_009047.
FT 2655 Missing (in isoform 3).
FT 1762 /FTid=VSP_009049.
FT 1831 Missing (in isoform 4).
FT 2592 /FTid=VSP_009050.
FT AISCIGPKAPFTNGGILTDYLVGRVTVFCNDGVRLSKEL
FT TTAVCQ -> GEVYAKNNKNNVRLAPFNVITFNSEN
FT GNIRKHIVNSPHKNA (in isoform 5).
FT /FTid=VSP_009051.
FT 3670 Missing (in isoform 5).
FT 182 /FTid=VSP_009052.
FT I -> M (in dbSNP:2219898).
FT /FTid=VAR_017404.
```

```
Query Match 3.0%; Score 106; DB 1; Length 3670;
Best Local Similarity 19.0%; Pred. No. 63;
Matches 142; Conservative 92; Mismatches 253; Indels 262; Gaps 40;

Qy 2 LCHAADTTGSPMPDTCVLLTSEBQAKWRQLEGEKSLGPHSPETPYI-----52
Db 197 ICRAEDACGTMRGSGI---ISSPSFPNHNADCTWTIVABPGDTISLIFTDFOMEE 253
Qy 53 KY---LEGSE-----TWKVKVLPD-----GISA-----73
Db 254 KYDYLEIEGSEPTIWSGMNIPPIISKNKWLHFVTDNSNHYRGFSAPYQSSLTILTH 313
Qy 74 SKILGIMARVRIATALLAVVLAAPCLAFDEVTSAGVPEEHKHTGEGRHLQTCNDDAL 133
Db 314 TTSTGELEENRITTTGAIVASTPA---DVTSSVTAVTIHRLSEBQRQVQVTSRLNSGL 369
Qy 134 DPT-----APNRDNNVAFASR-RDAARRRDGTGTVCOITN---GETDLATWPHKSLPH 183
Db 370 DPNTSKDGLSPHPADTQSTRRRPRHABQIER--TKELAVVTHRVKKAIDFKSRGFKLPP- 426
Qy 184 DELGOVTADDFAILEDCLNGDFSI CEDVPAGDPA-GRLV-----NPTAAFAIDISGPA 236
Db 427 ---GKNSNKFSLNEGKITASNLCPD---PGEPEGKRI GSDPSLSTVQFSCDEYVL 481
Qy 237 FSATTIP-----PVPTLSSPELAAQL---AELYWMA 264
Db 482 QGAKSITCQRIAEVFAAWSDRHPVCKVKTCGSLNQLGPGSGTFTSPNPPFPQYDSNAQCVMVI 541
Qy 265 LARDV-PFMOYGTDEITTTAAANLGMGFPNLDVAVSIGSDGTV-DPFSQLFRATFVGVE 322
Db 542 TAVNTNKVIQINFEEFDLEI-----GYDTLTIGDGEVGDG-----RTVLQVL 584
Qy 323 TGPVPSQLLVNSFTIDAITVEPKQETPAPDLYNMVDPDEWLNIQNGGPPAGPELDEBLR 382
Db 585 TGSFV-----PDLI VSMSSOMLHLQT-----DESVCSVG 614
Qy 383 FINARDLARVSFVDNINTEAYRGSLLLELGAFSRPGINGPFIDSDRQAGFVN-----F 437
Db 615 FKVNYEIEKESCGD-----PGT--PLYGIREGDFGFSNRDLRF 651
Qy 438 GTSHYFLIGAAELAQRASYQKQVH-----RPAR-----PEALGGTLH 477
Db 652 ECQFGFELIGEKSIVCOEN--NQWSANIPICIPCLSNFTAPMGTVLSPDYPEGYGNLNL 709
Qy 478 N--TIAGDLADPDLSLENDLLELLKRVAEINAAQNPNNVEVYLLPOAIQVGSPTHPSPYS 535
Db 710 CIWTIISDPCSRHLSP--NDFLESQDFLAVKGDGS-----PESPILGTTFTGAEVPS 761
Qy 536 GHATONGAFATVLKALIGLD-----RG-----GECFPNPVFPSP-----DGLLEL- 574
Db 762 -HLTSN---SHILRLBPOADHSMGRGFNITYNTFGHNEC-PDFGIPINARRFGDNFQLG 816
Qy 575 -----INPEG-----ACLYTEGEI 588
Db 817 SSISVICEEGFIKTQGTETITCILMDGKV 845

RESULT 11
SECD_RHOCA STANDARD; PRT; 554 AA.
ID SECD_RHOCA AC O33517;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Protein-export membrane protein secd.
GN SECD.
OS Rhodobacter capsulatus (Rhodopseudomonas capsulata).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC Rhodobacteraceae; Rhodobacter.
OX NCBI_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SB1003 / St Louis;
```

```
RX MEDLINE=97318920; PubMed=9175857;
RA Goldman B.S., Beckman D.L., Bali A., Monika E.M., Gabbert K.K.,
RA Kranz R.G.;
RT "Molecular and immunological analysis of an ABC transporter complex
RT required for cytochrome c biogenesis.";
RL J. Mol. Biol. 268:724-738 (1997).
CC -1- FUNCTION: Involved in protein export (By similarity).
CC -1- SUBUNIT: Part of the prokaryotic protein translocation apparatus
CC which comprise secA, secB, secD, secE, secF, secG and secY (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -1- SIMILARITY: Belongs to the secD/secF family. SecD family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U69979; AAB62801.1; -.
CC InterPro; IPR005791; SecD.
CC InterPro; IPR003335; SecD_SecF.
CC Pfam; PF02355; SecD_SecF; 1.
CC TIGRFAMs; TIGR00916; 2A0604s01; 1.
CC TIGRFAMs; TIGR01129; secD; 1.
CC
CC Protein transport; Translocation; Transmembrane; Membrane.
FT TRANSMEM 10 30 POTENTIAL.
FT TRANSMEM 392 412 POTENTIAL.
FT TRANSMEM 435 455 POTENTIAL.
FT TRANSMEM 491 511 POTENTIAL.
FT TRANSMEM 516 536 POTENTIAL.
SQ SEQUENCE 554 AA; 58943 MW; DF2CBEBA9F69EDF CRC64;
Query Match 3.0%; Score 104.5; DB 1; Length 554;
Best Local Similarity 23.5%; Pred. No. 5.4;
Matches 96; Conservative 48; Mismatches 169; Indels 95; Gaps 16;
QY 79 KIMARVRIATALAVLAAPCLAFDEVTASGVFPPEHKHGTGEG-----RHLOTCTNSDD 131
Db 128 QIGERAQIARAVEV---ARTLSPVSLTGVGTQDYEVTGEGTVVFLSEAKKATDDR 184
QY 132 ALDPTAPNRDNNVAFSRD-----AARERDGTGTVCQIT----- 167
Db 185 TMOQSLIEVRRVDAAGTRPTIMREGTRILIEVPGISGAQLKDLIGTAKLTTHPVL 244
QY 168 -----NGETLATWFKSLPH--DELQGVTDADDFAILEDCLNCD----- 205
Db 245 STTSNPNAPVAGSNELLPDAERQGLYHLLDVPVVTGDD---LTDARPTTDDGAPAVSF 301
QY 206 -FSICEDVPAGD-PAGRLVNPTAAFAIDISGPAFSATTIPP-----VPTLSSPELAA 255
Db 302 RNVSGARAFGDTYAGHGEF---FAVLGKVISAPTIQAHGASGITGFTSEAT 358
QY 256 QLAELVYM-ALARDVPPM-----QYGTDEITTTAAANLAGMGGFNLDAVSGSDGTV 307
Db 359 DLALLRAGALPAGMTFLEERTIGPELGADSV---KAGWASVIGFVAVVYMIASVGLF 415
QY 308 DPFSQLFRATFVGVEGPFVQLVNSFTTDAITVEPKQTFPAPDLNMYVDFEWNINQ 367
Db 416 GFPESSV--ALFINIAFVYMGVAGTGTMLPGIAGIVLLTGTSDANVLIERMREIRS 473
QY 368 GGPAPGPELDELRLFRNRDARLVSVFDVNTINTEAVRGLLLELGA 415
Db 474 GKSP-----VRAIELGFDKAMSAIID-ANVTSFLSAILFLVGA 511
RESULT 12
MUTA MYCTU
ID MUTA MYCTU
AC P71773; STANDARD; PRT; 615 AA.
DT 01-NOV-1997 (Rel. 35, Created)
```

```
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable methylmalonyl-CoA mutase small subunit (EC 5.4.99.2) (MCM).
GN MUTA OR RV1492 OR MT1539 OR MTCY277.14 OR MB1529.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellton S., Squares R., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
CC -1- FUNCTION: Catalyzes the isomerization of succinyl-CoA to
CC methylmalonyl-CoA during synthesis of propionate from
CC tri-carboxylic acid-cycle intermediates (By similarity).
CC -1- CATALYTIC ACTIVITY: (R)-2-methyl-3-oxopropionyl-CoA = succinyl-
CC CoA.
CC -1- COFACTOR: Adenosylcobalamin (By similarity).
CC -1- PATHWAY: Protonic acid fermentation.
CC -1- SUBUNIT: Heterodimer of an alpha and a beta chain (By similarity).
CC -1- SIMILARITY: Belongs to the methylmalonyl-CoA mutase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z79701; CAB02042.1; -.
CC EMBL; A5007022; AKA45806.1; -.
CC EMBL; BX248339; CAD96196.1; -.
CC PIR; G70711; G70711.
CC HSSP; P11652; 2REQ.
CC TIGR; MT1539; -.
CC TubercuList; Rv1492; -.
CC InterPro; IPR006099; MMCoA_mutase.
CC InterPro; IPR004608; MMCoA_mutase_b.
```



Db 476 PAFTATSSSQIKQIGPELPRP-----NPVRIPTRRRELASFGIKLP 518  
Qy 152 AAR-----RRDGTGTVCQITNGETDLATMFHKSPLHDELQGVTTADDFAILEDCLINGD 205  
Db 519 SQWAEQELRERG-----DETQNPQWAASSYG-----TEITSDEDAALQAILRKA 565  
Qy 206 FSCIEDVPAGDPAGRLVNPFAAFAIDIS--GPAFSAFTTIPVPTLSSPELAQAELLYMM 263  
Db 566 F-----ADQOSRYALSTLAEOSSITERSPAAEMPTTTSQVSDLEDEQALQAEALRQA 618  
Qy 264 ALARDVPMQVGTDEITTTAAANLAGMG--GFNLDVAVSGISGTVDPSPQLPRATFVG 320  
Db 619 FAQQQ--HRYGAGTGDNDVAVNIRSDVTSTFTSPIADLVDSPREPLFTL----- 669  
Qy 321 VETGPFVSQLVNSFTDAITVEPKQETFAPD 352  
Db 670 ---SPYVDETDV-----EPVQLECKERSLLQD 694

RESULT 14  
COMP\_MOUSE  
ID\_COMP\_MOUSE STANDARD; PRT; 755 AA.  
AC QSR0G6; 2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Cartilage oligomeric matrix protein precursor (COMP).  
GN COMP.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Cartilage;  
RX MEDLINE=20505681; PubMed=11052496;  
RA Fang C., Carlson C.S., Leslie M.P., Tulli H., Stolerman E., Perris R.,  
RA Ni L., Di Cesare P.E.;  
RT "Molecular cloning, sequencing, and tissue and developmental  
RT expression of mouse cartilage oligomeric matrix protein (COMP).";  
RL J. Orthop. Res. 18:593-603(2000).  
RN [2]  
RX X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 28-71.  
RX MEDLINE=22313308; PubMed=12426368;  
RA Ozbek S., Engel J., Stetefeld J.;  
RT "Storage function of the coiled-coil domain in complex with vitamin D(3).";  
RT structure of the coiled-coil domain in complex with vitamin D(3).";  
RL EMBO J. 21:5960-5968(2002).  
CC -!- SUBUNIT: Pentamer; disulfide-linked (By similarity).  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- SIMILARITY: Belongs to the thrombospondin family.  
CC -!- SIMILARITY: Contains 4 EGF-like domains.  
CC -!- SIMILARITY: Contains 7 TSP type-3 domains.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AF033530; AAD01972.1; --  
DB; 1M29; 28-JAN-03.  
DR MGD; MGI:88469; Comp.  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR001881; EGF Ca.  
DR InterPro; IPR006209; EGF-like.  
DR InterPro; IPR003367; tsp\_3.  
DR InterPro; IPR008859; TSPC.  
DR Pfam; PF00008; EGF; 1.  
DR Pfam; PF05735; TSPC; 1.  
DR Pfam; PF02412; tsp\_3; 13.

DR SMART; SM00179; EGF CA; 2.  
DR PROSITE; PS00018; EF\_HAND; 1.  
DR PROSITE; PS00022; EGF\_1; FALSE\_NEG.  
DR PROSITE; PS01186; EGF\_2; 1.  
DR PROSITE; PS00026; EGF\_3; 3.  
DR PROSITE; PS01187; EGF\_CA; 2.  
KW Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain;  
KW Signal; 3D-structure.  
FT SIGNAL 1 19  
FT CHAIN 20 755  
FT DOMAIN 21 84  
FT N-TERMINAL.  
FT DOMAIN 85 124  
FT EGF-LIKE 1.  
FT DOMAIN 125 177  
FT EGF-LIKE 2. CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 178 220  
FT EGF-LIKE 3. CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 223 265  
FT EGF-LIKE 4.  
FT DOMAIN 295 330  
FT TSP TYPE-3 1.  
FT DOMAIN 331 353  
FT TSP TYPE-3 2.  
FT DOMAIN 354 389  
FT TSP TYPE-3 3.  
FT DOMAIN 390 412  
FT TSP TYPE-3 4.  
FT DOMAIN 413 450  
FT TSP TYPE-3 5.  
FT DOMAIN 451 486  
FT TSP TYPE-3 6.  
FT DOMAIN 487 522  
FT TSP TYPE-3 7.  
FT C-TERMINAL.  
FT DOMAIN 523 755  
FT INTERCHAIN.  
FT DISULFID 68 68  
FT DISULFID 71 71  
FT DISULFID 89 100  
FT BY SIMILARITY.  
FT DISULFID 94 109  
FT BY SIMILARITY.  
FT DISULFID 112 123  
FT BY SIMILARITY.  
FT DISULFID 129 140  
FT BY SIMILARITY.  
FT DISULFID 134 149  
FT BY SIMILARITY.  
FT DISULFID 152 176  
FT BY SIMILARITY.  
FT DISULFID 182 195  
FT BY SIMILARITY.  
FT DISULFID 189 204  
FT BY SIMILARITY.  
FT DISULFID 207 219  
FT BY SIMILARITY.  
FT DISULFID 227 241  
FT BY SIMILARITY.  
FT DISULFID 235 251  
FT BY SIMILARITY.  
FT DISULFID 253 264  
FT BY SIMILARITY.  
FT DISULFID 280 285  
FT BY SIMILARITY.  
FT DISULFID 290 310  
FT BY SIMILARITY.  
FT DISULFID 326 346  
FT BY SIMILARITY.  
FT DISULFID 349 369  
FT BY SIMILARITY.  
FT DISULFID 385 405  
FT BY SIMILARITY.  
FT DISULFID 408 428  
FT BY SIMILARITY.  
FT DISULFID 446 466  
FT BY SIMILARITY.  
FT DISULFID 482 502  
FT BY SIMILARITY.  
FT DISULFID 518 739  
FT BY SIMILARITY.  
FT CARBOHYD 119 119  
FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 740 740  
FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 755 AA; 82352 MW; 7DDFCF443589A0B7 CRC64;

Query Match 2.9%; Score 103.5; DB 1; Length 755;  
Best Local Similarity 24.2%; Pred. No. 9.9;  
Matches 51; Conservative 19; Mismatches 90; Indels 51; Gaps 10;  
Qy 112 EEKHGTGEGRHLQCTNSDDALDPTAPNRRDNNVAFSRDRAARRDGTGTVCQITNGET 171  
Db 455 QDSHDCKG--DACDDDD--NDGVDSRDNCLVNPQGDNDRDGVGDAQ--GDF 506  
Qy 172 DLATMFHKSPLHDELQGVTTADDFAILEDCLIN--GDFS-----CEDVA--GDP 217  
Db 507 DADKVIDKIDVCPENAEVTLTDFRAFQTVVLDPEGDAQIDPNWVVLNQGMIEIVQTMNSDP 566  
Qy 218 AGRLVNPTAAFAIDISGPFAFSATTIPVPTLSSPELAA-----QLAEL 260  
Db 567 -GLAVGYTAFNGVDPEGTFH-----VNTATDDYAGPIFGYQDSSSFYVMWKMEQT 618  
Qy 261 YWMALARDVPFMQVGTDEITTTAAANLAGMG 291  
Db 619 YWQA----NPFRAVAEPGIQLKAVKSTGPG 645

RESULT 15  
COMP\_RAT

